0830-410
0312
NASA.55:3095
JAN 12 1991

NASA Conference Publication 3095

Joint University Program for Air Transportation Research 1989–1990

Proceedings of a conference held in
Athens, Ohio
June 14–15, 1990

NASA

NASA Conference Publication 3095

Joint University Program for Air Transportation Research 1989–1990

Compiled by
Frederick R. Morrell
Langley Research Center
Hampton, Virginia

Proceedings of a conference sponsored by the National Aeronautics and Space Administration and the Federal Aviation Administration and held in Athens, Ohio
June 14–15, 1990

NASA
National Aeronautics and Space Administration
Office of Management
Scientific and Technical Information Division
1990

I

PREFACE

The Joint University Program for Air Transportation Research is a coordinated set of three grants sponsored by NASA Langley Research Center and the Federal Aviation Administration, one each with the Massachusetts Institute of Technology (NGL-22-009-640), Ohio University (NGR-36-009-017), and Princeton University (NGL-31-001-252). These research grants, which were instituted in 1971, build on the strengths of each institution. The goals of this program are consistent with the aeronautical interests of both NASA and the FAA in furthering the safety and efficiency of the National Airspace System. The continued development of the National Airspace System, however, requires advanced technology from a variety of disciplines, especially in the areas of computer science, guidance and control theory and practice, aircraft performance, flight dynamics, and applied experimental psychology. The Joint University Program was created to provide new methods for interdisciplinary education to develop research workers to solve these large scale problems. Each university submits a separate proposal yearly and is dealt with individually by NASA and the FAA. At the completion of each research task, a comprehensive and detailed report is issued for distribution to the program participants. Typically, this is a thesis that fulfills the requirements for an advanced degree or a report describing an undergraduate research project. Also, papers are submitted to technical conferences and archival journals. These papers serve the Joint University Program as visibility to national and international audiences.

To promote technical interchange among the students, periodic reviews are held at the schools and at a NASA or FAA facility. The 1989–1990 year-end review was held at Ohio University, Athens, Ohio, June 14–15, 1990. At these reviews the program participants, both graduate and undergraduate, have an opportunity to present their research activities to their peers, to professors, and to invited guests from government and industry.

This conference publication represents the tenth in a series of yearly summaries of the program. (The 1988/89 summary appears in NASA CP-3063). Most of the material is the effort of students supported by the research grants. Four types of contributions are included in this publication: a summary of ongoing research relevant to the Joint University Program is presented by each principal investigator, completed works are represented by full technical papers, research previously in the open literature (e.g., theses or journal articles) is presented in an annotated bibliography, and status reports of ongoing research are represented by copies of presentations with accompanying text.

Use of trade names of manufacturers in this report does not constitute an official endorsement of such products or manufacturers, either expressed or implied, by the National Aeronautics and Space Administration or the Federal Aviation Administration.

Frederick R. Morrell
NASA Langley Research Center

CONTENTS

PREFACE iii

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 1989–1990 3
Robert W. Simpson

AUTOMATIC SPEECH RECOGNITION IN AIR TRAFFIC CONTROL: A HUMAN FACTORS PERSPECTIVE 9
Joakim Karlsson

THE TEMPORAL LOGIC OF THE TOWER CHIEF SYSTEM 15
Lyman R. Hazelton, Jr.

BUBBLES—AN AUTOMATED DECISION SUPPORT SYSTEM FOR FINAL APPROACH CONTROLLERS 29
Zhizang Chi

ANALYSIS OF AIRCRAFT PERFORMANCE DURING LATERAL MANEUVERING FOR MICROBURST AVOIDANCE 37
Denise Ávila De Melo and R. John Hansman, Jr.

HAZARD EVALUATION AND OPERATIONAL COCKPIT DISPLAY OF GROUND-MEASURED WINDSHEAR DATA 45
Craig Wanke and R. John Hansman, Jr.

OHIO UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT OHIO UNIVERSITY, 1989–1990 61
Robert W. Lilley

RIDGE REGRESSION SIGNAL PROCESSING 67
Mark R. Kuhl

MODELING SELECTIVE AVAILABILITY OF THE NAVSTAR GLOBAL POSITIONING SYSTEM 77
Michael Braasch

ANALYZING THOUGHT-RELATED ELECTROENCEPHALOGRAPHIC DATA USING NONLINEAR TECHNIQUES 91
Trent Skidmore

INTEGRATED INERTIAL/GPS 97
Paul Kline and Frank van Graas

PRINCETON UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT PRINCETON UNIVERSITY, 1989–1990 105
Robert F. Stengel

PROGRESS ON INTELLIGENT GUIDANCE AND CONTROL FOR WIND SHEAR ENCOUNTER 119
D. Alexander Stratton

FLIGHT SIMULATION FOR WIND SHEAR ENCOUNTER 133
Sandeep S. Mulgund

STOCHASTIC ROBUSTNESS: TOWARDS A COMPREHENSIVE ROBUSTNESS TOOL 141
Laurie Ryan Ray

NEURAL NETWORKS IN NONLINEAR AIRCRAFT CONTROL 151
Dennis J. Linse

STABILITY BOUNDARIES FOR AIRCRAFT WITH UNSTABLE LATERAL-DIRECTIONAL DYNAMICS AND CONTROL SATURATION 163
Prakash C. Shrivastava and Robert F. Stengel

RESTRUCTURABLE CONTROL USING PROPORTIONAL-INTEGRAL IMPLICIT MODEL FOLLOWING 173
Chien Y. Huang and Robert F. Stengel

STOCHASTIC STABILITY AND PERFORMANCE ROBUSTNESS OF LINEAR MULTIVARIABLE SYSTEMS 181
Laura E. Ryan and Robert F. Stengel

SYSTEM IDENTIFICATION FOR NONLINEAR CONTROL USING NEURAL NETWORKS 187
Robert F. Stengel and Dennis J. Linse

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY 1989-1990

Robert W. Simpson
Flight Transportation Laboratory
Massachusetts Institute of Technology
Cambridge, MA

1. INTRODUCTION

There are one completed project and four areas of research under way under the sponsorship of the FAA/NASA Joint University Research Program during 1990.

The completed project was on "The Integration of Automatic Speech Recognition into the ATC System", FTL Report R90-1, January 1990 (see Reference 1). A brief paper on this work is presented here.

The four active research projects are:

- An Expert System for Runway Configuration Planning
- An Automated Decision Support System for Final Approach Spacing
- Modeling of Ice Accretion on Aircraft in Glaze Icing Conditions
- Cockpit Display of Hazardous Weather Information

2. AN EXPERT SYSTEM FOR RUNWAY CONFIGURATION PLANNING

The research effort in this area has been exploring the application of expert systems technology to the problem of creating a daily schedule for the operation of runways at major airports of the USA. There are dozens of runway operating configurations when the assignment of aircraft by type, by landing or take-off, under wet or dry surface conditions, etc., are used to define a configuration, and since they can have a wide range of landing and take-off capacities, it is necessary to produce a forecast of the runway operations in order to provide a capacity forecast to the ATC* Central Flow Control Facility. There are a wide variety of rules which constrain the daily schedule, and a set of conflicting objectives: windspeed and direction may

*Air Traffic Control

eliminate some runways; visibility and ceiling may eliminate some ATC procedures to feed the runways; surface conditions may stop other ATC practices; there may be a need to plan snow removal or runway maintenance; there is a desire to minimize delays; there is a desire to equalize noise on communities and to grant relief after a few hours of operations on any runway end; there is a desire to avoid frequent changes, and to avoid scheduling a change when the ATC controller shift is changing.

Given these rules and the knowledge that they will change over time and are different at each airport, it was decided to explore the use of Expert Systems software technology as a method of creating a flexible, adaptable decision support tool for ATC supervisors. However, it has been found to be sadly deficient in facing up to real-time, changing inputs and in maintaining a consistent, explainable plan. The challenges of the requirements of this problem have caused interesting basic research of the software technology for Expert Planning Systems to be conducted. It now appears that an elegant method of maintaining the relationships for the "truths" of an Expert System when the facts are changing over time has been discovered, and that a method to explain not only the reasons for the current plan can be provided, but also an historical explanation of the reasons for the time history of plans. Thus, an Expert System for Temporal Planning has been created. A brief exposition on certain aspects of temporal logic appears here as a brief annotated presentation by Lyman Hazelton.

3. AN AUTOMATED DECISION SUPPORT SYSTEM FOR FINAL APPROACH SPACING

Research into this area was initiated during 1990. It is interested in using improved display graphics to provide an interactive cueing system which assists the Final Approach Controller at a busy airport. It assumes the following:

1) An explicit schedule for landings and takeoffs is being constantly produced for all runway operations.

2) Each landing aircraft has declared an IAS* for final approach and is obligated to fly that approach speed.

3) There is a continuous updating of average windspeed on final approach.

The objective is to create an interactive system for providing cues to spacing commands to landing aircraft by the final approach controller which has the following attributes:

*Indicated Air Speed

a) The cues are adaptive to estimation errors in position and speed derived from a typical radar tracking process, and to the piloting errors in execution of turns and commanded speed reductions.

b) The cues are responsive to the desires of the ATC controller who remains in charge of the traffic situation and who can change the landing sequence, insert a missed approach aircraft into the sequence, increase or decrease spacing for any particular landing aircraft, and can see the planned insertions of take-offs, etc.

A simple, robust method of creating spacing cues for "Turn to Base", "Turn to Intercept", and "Reduce to Final Approach Speed" has been coded. A method of modifying the desired landing schedule by mousing and moving "schedule bubbles" on the extended runway centerline has been developed. Research is under way to create a self-adaptive correction to the runway schedule due to final errors in achieving the desired spacing, such that an aircraft which is "out-of-schedule" on final approach will cause a modification to the subsequent schedule and its cues. Further plans involve showing the inserted take-offs on the extended runway centerline upon request, and making them reactive as the controller shifts landing schedule bubbles; and then to impose a logic which maximizes total runway operations by ensuring that landing spacings do not preclude a possible take-off insertion. This is accomplished by automatically shifting the scheduled landing later by some small amount. An annotated set of viewgraphs by Robert W. Simpson is shown in this report. The student researcher is Zhihang Chi.

4. MODELING OF ICE ACCRETION ON AIRCRAFT IN GLAZE ICING CONDITIONS

The work in aircraft icing over the past year has continued to extend the multi-zone glaze ice accretion model which had been developed previously. The multi-zone model recognizes distinct zones of surface roughness and corrects the convective heat transfer for surface roughness effects. Since glaze ice accretion is dominated by heat transfer processes, this has a major effect on the shape of the accreted ice. The work this year has focused on understanding correlation between surface roughness and heat transfer and understanding what causes the observed changes in surface roughness. The surface roughness transition has been linked with the laminar to turbulent boundary layer transition. This understanding has lead to an improved version of the multi-zone model. Significant improvements in glaze ice accretion modeling have been accomplished with this model. A paper presented by Keiko Yamaguchi and R. John Hansman (Ref. 2) is included in this report.

5. COCKPIT DISPLAY AND EVALUATION OF HAZARDOUS WEATHER ALERTS

The work on hazardous weather alerts has continued to focus on microburst alerting. Various modes (voice, alphanumeric, and graphical) of information transfer from ground based sensors have been evaluated in part task simulator studies. The influence of geometrical, measurement, and procedural issues on Doppler Weather Radar hazard threshold criteria have been studied. In addition, a performance analysis has been done to study the advantages and disadvantages of lateral maneuvering for microburst avoidance. Papers presented by Denise Avila de Melo (Ref. 3) and Craig Wanke (Ref. 4) both with R. John Hansman are included in this report.

6. ANNOTATED REFERENCES OF 1990 PUBLICATIONS

1. Karlsson, Joakim, The Integration of Automatic Speech Recognition into the ATC System, FTL Report R90-1, January 1990, Flight Transportation Laboratory, MIT, Cambridge, MA 02139.

Today, the Air Traffic Control (ATC) system relies primarily on verbal communication between the air traffic controllers and the pilots of the aircraft in the controlled airspace. Although a computer system exists that processes primary radar, secondary radar, and flight plan information, the information contained within the verbal communications is not retained. The introduction of Automatic Speech Recognition (ASR) technology would allow this information to be captured for processing.

The research presented in this paper examines the feasibility of using ASR technology in the Air Traffic Control environment. The current status of the technology is assessed. Problems that are unique to ATC applications of voice input are identified. Since ASR technology is inherently a part of the man-machine interface between the user and the system, emphasis is placed on the relevant human factors issues. A man-machine model is presented which demonstrates the use of mixed input modalities, automatic error detection and correction techniques, and the optimal use of feedback to the controller.

Much of the potential benefit of introducing ASR technology into the Air Traffic Control system is a result of the highly constrained language used by air traffic controllers. Consequently, the information content of the ATC language must be determined, and methods must be designed to process the various levels of knowledge inherently available in ATC communications. The man-machine model adopted in this paper demonstrates techniques to utilize syntactic, semantic, and pragmatic information to improve overall recognition accuracy. An intelligent, adaptive voice input parser is presented.

2. Yamaguchi, K., and Hansman, R.J., "Heat Transfer on Accreting Ice Surfaces," AIAA-90-0200, AIAA 28th Aerospace Sciences Meeting, January 1990.

The influence and modeling of surface roughness for the numerical simulation of glaze ice accretion is studied. Icing wind tunnel tests are discussed and correlations between experimental and predicted heat transfer coefficients are presented.

3. Avila de Melo, D. and Hansman, R.J., "Analysis of Aircraft Performance During Lateral Maneuvering for Microburst Avoidance," AIAA-90-0568, AIAA 28th Aerospace Sciences Meeting, January 1990.

An analysis is conducted of a B-737-100 using a nonlinear aircraft dynamic simulation to evaluate the performance advantages and disadvantages of lateral maneuvering for microburst avoidance.

4. Wanke, C., and Hansman, R.J., "Hazard Evaluation and Operational Cockpit Display of Ground Measured Windshear Data," AIAA-90-0566, AIAA 28th Aerospace Sciences Meeting, January 1990.

Information transfer issues and hazard evaluation issues associated with ground based doppler weather radar microburst alerts are studied. Part task simulator studies and numerical aircraft performance models are used.

AUTOMATIC SPEECH RECOGNITION IN AIR TRAFFIC CONTROL: A HUMAN FACTORS PERSPECTIVE*

Joakim Karlsson
Research Assistant
Flight Transportation Laboratory, M.I.T.
Cambridge, MA

ABSTRACT

The introduction of Automatic Speech Recognition (ASR) technology into the Air Traffic Control (ATC) system has the potential to improve overall system safety and efficiency. However, because ASR technology is inherently a part of the man-machine interface between the user and the system, the human factors issues involved must be addressed. This paper identifies some of the relevant human factors problems and presents related methods of investigation. Research at M.I.T.'s Flight Transportation Laboratory is being conducted from a human factors perspective, focusing on intelligent parser design, presentation of feedback, error correction strategy design, and optimal choice of input modalities.

INTRODUCTION

In today's ATC system, communication between controllers and aircraft is almost exclusively verbal. This is especially true for such critical tasks as the issuing of clearances and vectors, to achieve traffic separation. Although a digital datalink is in development (Mode S), there is no reason to believe that voice communication between ATC and aircraft will disappear in the near future. As a result, most of the information transferred within the system is never captured in machine readable form. Herein lies the promise of introducing ASR technology into the ATC system: it would permit processing of ATC clearances, to ensure conformance to safety and separation criteria. It would allow the ATC computer system to predict the future state of the airspace. The controller could prestore routine clearances during periods of little activity. Mode S equipped aircraft could be provided with a machine readable copy of verbal clearances for confirmation purposes.

Thus, introduction of ASR technology could result in the reduction of human errors, resulting in increased system safety. However, the dilemma of ASR is that its purported advantages are not automatically realized by simply making the technology available. Careful human factors design is necessary to capitalize on its potential [Berman, 1984]. This is especially true in the case of ATC, which is plagued by human factors problems such as intense levels of workload during traffic peaks intermixed with controller boredom during low demand periods. Furthermore, the high probability of loss of lives in the case of errors makes it imperative that the human factors problems created by introducing ASR into the Air Traffic Control system are properly addressed and solved.

The speech recognition devices available today are not sufficiently capable to be used operationally within the ATC environment. However, there are units available that are useful for the required preliminary human factors research. In order to minimize human factors problems, it is necessary to implement an iterative design cycle that should be continued until the needs of the system users are met [Cooper, 1987]. The research presented within this paper should be considered as one step in that cycle.

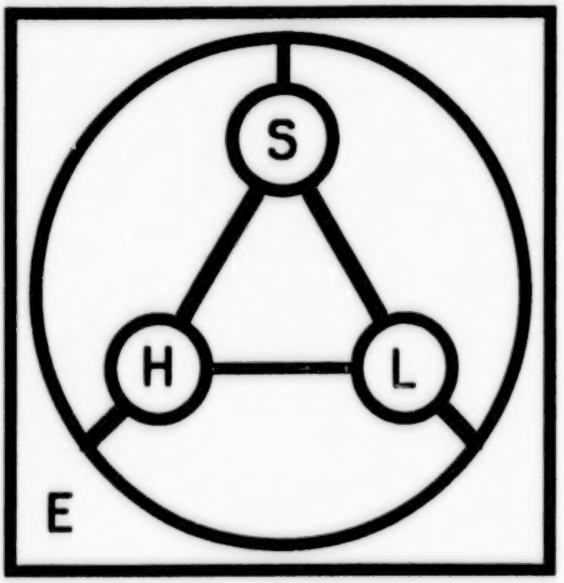
---

*Paper presented at Military and Government Speech Technology 1989, Nov. 13–15, 1989, Arlington, VA.

MODELING HUMAN FACTORS

In order to approach human factors in an analytic way, a conceptual model of the system resources available can be used. The system resources include hardware, software (rules and regulations), liveware (users), and the environment. The SHEL model, named by the initial letters of these resources, can be used to represent the components and their links [Edwards, 1988]. Figure 1 contains a graphical representation of the SHEL model. The connecting lines between the system components represent the interfaces between the respective resources. It is at the interfaces to the liveware component that most human factors issues occur.

Figure 1: The SHEL Model.



In terms of the SHEL model, examples of human factors problems include microphone placement and characteristics (hardware-liveware interface), speech variations due to background noise (environment-liveware interface), and design of error correction strategies (software-liveware interface). Note that not all human factors issues are strictly related to one single interface to the liveware component. Examples include fatigue, stress, boredom, and user acceptance of ASR technology. It should also be emphasized that ATC is a multi-user system. Thus, there are also liveware-liveware interfaces that must be considered.

IDENTIFYING HUMAN FACTOR ISSUES

Identifying human factors issues related to ASR technology is a topic that has been covered adequately and extensively [Constantine, 1984]. However, ATC is fundamentally different from other ASR applications in several ways:

- In ATC, voice is the primary communication channel, and microphones are already used.
- The ATC vocabulary and syntax are already defined and cannot be easily altered.
- Human errors in the ATC environment can lead to fatal results.
- The background noise consists of distinct voices, not random noise.

Hence, we can consider three categories of human factors issues: common issues that are mutual to both ATC and other ASR applications, unique issues that are typically not encountered in other applications, and non-issues – problems that may be significant in other applications, but that do not play a major role in ATC.

The last group, non-issues, is of course the most trivial to consider: a good example is the hardware-liveware issue of microphone characteristics and placement. Headset mounted noise-cancelling microphones are already in use in the current ATC environment, and hence it is an issue that has been addressed extensively before ASR technology has come under consideration. Also, the software-liveware problem of vocabulary and syntax definition, normally an important human factors issue, has also been completed: these definitions are controlled by the Federal Aviation Administration (FAA). Another common problem, communication with other people while in recognition mode, is not likely to occur in ATC, as controllers are already using Push-To-Talk (PTT) switches on their headsets.

Issues that are common to both ATC and other ASR applications are abundant, and must not be neglected, although they have already been addressed extensively. These include:

- Speech variations due to stress, fatigue, or background noise.
- Spurious recognition due to background noise.
- Inter-speaker variations (the "sheep and goats" issue).
- User acceptance of the technology.
- User motivation.
- Presentation of feedback to the user.
- Error recognition, presentation, and correction.
- User training.
- Selection of proper hardware.
- Optimal use of mixed input modalities.
- Recognition accuracy and use of higher levels of knowledge.
- Failure to adhere to the syntax.

Although several of these problems remain unsolved, most have been addressed previously. The research being conducted at the Flight Transportation Laboratory covers some of these issues, since much of the previous work has not been applied specifically to ATC.

The final group, unique issues, includes problems that are either specific to ATC, or that are more significant in the ATC environment than elsewhere. A typical example is stress induced reduction of recognition accuracy, mentioned above as an issue common to other ASR applications. This is a much more critical issue in Air Traffic Control, since the cases where the introduction of ASR has the greatest potential of improving safety, are likely to be stressful situations. The possibility of automating conformance monitoring would greatly benefit the controller during scenarios where a large number of aircraft are being controlled – a stressful period for the controller. It is exactly in the conditions where ASR technology is needed most, that it performs worst. To the human factors researcher this points out the importance of high baseline recognition accuracy, high levels of robustness in the presence of speech variations, introduction of functional automatic error correction techniques, and the design and implementation of parsers that make use of higher levels of knowledge such as prosodic, syntactic, semantic, and pragmatic information.

Another major issue facing the introduction of ASR technology into the ATC environment is that of cognitive workload. The controllers are already presented with a wealth of information, and if any new technology is to be introduced it must reduce workload, not increase it. There exists a need to ensure that the information captured through the use of speech input technology is the same as the information transmitted to the aircraft. Hence, the controller must monitor what is understood by the machine, in order to be able to correct it. However, this would introduce another task for the controller, and possibly distract from the visual attention that the radar display demands. This dilemma underscores the importance of designing adequate feedback and error correction strategies.

RESEARCH AT THE M.I.T. FLIGHT TRANSPORTATION LABORATORY

It is within the framework presented above that the ASR research effort at M.I.T.'s Flight Transportation Laboratory has been conducted. Only a brief description of this research can be presented in this paper – more detailed descriptions are available elsewhere [Karlsson, 1990]. Preliminary results include a study to choose the ASR hardware most suited for ATC human factors research, purchase and evaluation of the Votan VPC 2000 and Verbex Series 5000 voice I/O systems, and design and implementation of a low-cost portable research station using the Verbex Series 5000 and a PC based ATC simulator. An extensive annotated bibliography of related papers has also been compiled.

Future work will concentrate on means to improve recognition accuracy while maintaining a low workload level. Techniques will include the use of semantic and pragmatic information, adaptive (on-the-fly) training, introduction of confusability matrices and other automatic error correction techniques [Loken-Kim, 1985], mouse and menu input for error correction, and need-to-know type feedback that ensures that the use of ASR technology remains mostly transparent to the controller. Furthermore, a receiver station has been established to monitor ATC communication in the greater Boston area, to study real life use of the ATC language and provide data for issues such as syntax deviation.

CONCLUSIONS

The importance of the human factors aspects of introducing ASR technology into the ATC environment cannot be underestimated. In particular, it must be realized that ATC applications are uniquely different from other applications where voice input may be of benefit. As a result, much greater emphasis must be placed on issues such as mental workload, user feedback, mixed use of input modalities, intelligent parser design, and improved robustness with respect to speech variations. The ASR research being conducted at the M.I.T. Flight Transportation Laboratory has resulted in a set of tools that can be used to identify, quantify, and provide preliminary solutions to the human factors issues described within this paper. The results can then be used as a step in an iterative design cycle to obtain a system acceptable to the user.

REFERENCES

Berman, J. V. F., "Speech Technology in a High Workload Environment", *Proceedings of the 1st International Conference on Speech Technology*, IFS Publications Ltd, Bedford, U.K. (1984), pp. 69-76.

Constantine, Betsy J., "Human Factors Considerations in the Design of Voice Input/Output Applications", *The Official Proceedings of Speech Tech '84*, Media Dimensions, Inc., New York, N.Y. (1984), pp. 219-224.

Cooper, Martin, "Human Factors Aspects of Voice Input/Output", *Speech Technology*, Vol. 3, No. 4 (March/April 1987), pp. 82-86.

Edwards, Elwyn, "Introductory Overview", *Human Factors in Aviation*, eds Wiener, Earl L. and Nagel, David C., Academic Press, Inc., San Diego, CA (1988), pp. 3-25.

Karlsson, Joakim, "Automatic Speech Recognition in Air Traffic Control", *Joint University Program for Air Transportation Research – 1988,1989* NASA Conference Publication 3063, NASA Langley Research Center, Hampton, VA (1990), pp. 3-15.

Karlsson, Joakim, *The Introduction of Automatic Speech Recognition Technology into the Air Traffic Control System*, FTL Report 90-1,Flight Transportation Laboratory, M.I.T., Cambridge, MA (Jan. 1990).

Loken-Kim, K.H., "An Investigation of Automatic Error Detection and Correction in Speech Recognition Systems", *The Official Proceedings of Speech Tech '85*, Media Dimensions, Inc., New York, N.Y. (1985), pp. 72-74.

The Temporal Logic of the Tower Chief System

Lyman R. Hazelton, Jr.*
Flight Transportation Laboratory
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA

1 Introduction

Human reasoners engage in diverse and powerful forms of cognition. Perhaps the most common mechanism used by humans is induction, which is the usual process of learning. The most celebrated and well understood process of thought is deduction, the method employed in mathematical proof. Humans can also make plans, reason about events and causality in an environment which frequently changes and often provides incomplete information. Humans need not be certain about their facts, and can make assumptions when required information is unavailable.

Classical logic deals almost exclusively with deduction. In chapter 4 of Aristotle's Metaphysics [1], there is a section regarding the application of logic to predicting the future. In that treatise, Aristotle describes what has now become known as the axiom of Excluded Middle. In essence, he states that in his logic a proposition can be either TRUE or FALSE, exclusively. No other values for the veracity of a proposition are possible. There is no middle ground. Because the future is indeterminate, one cannot assign absolute knowledge of a forthcoming event. This is a broad statement which has far reaching implications. As pointed out by Bertrand Russell in the early nineteen hundreds, it is not a trifling thing to easily toss aside.

On the other hand, perhaps the interpretation of the axiom of excluded middle has been too broad. Aristotle was seeking truth in a very absolute way. He was quite aware of the works of Zeno and Pythagoras, and to a certain extent his logic was developed from his knowledge of their process of mathematical proof. Propositions in mathematics are *universal* in their temporal extent. One never hears a geometer state that two lines are parallel from two until four this afternoon. The lines are simply parallel or they are not. The geometer's proof makes no reference to time, and, in like manner, neither does Aristotle's logic.

One should not lose track of the fact that logic is primarily a model for analysis of human thought. Its utility as a paradigm for symbolic reasoning and computer programming is secondary to its value as a model of the cognitive process. The restriction to exclude time simplifies the model and makes it much more comprehensible. The restriction also limits the range of problems to which the model may be successfully applied.

Clearly, human reasoners often contemplate the future and create rational plans about it. During the last century, logicians have begun to extend classical logic in order to

*Research supported by NASA contract number NGL 22-009-640.

better model this kind of cognition. *Modal logics* have been invented to describe systems in which propositions may or might be true or false. *Defeasible logics* and *model theory* are used for systems which involve change. And *temporal logics* have been devised to reason about time. The purpose of this paper is to describe the logic used in the reasoning scheme employed in the *Tower Chief* system.

2 Classical Logic

Let us begin with a short review of the fundamental ideas of logic. The elementary operands of logic are propositions. A proposition is a declarative statement such as, "Socrates was a man", or "There is ice on runway 33." Logic consists of:

1. A set of abstract operations for combining propositions in ways that preserve the veracity of the resulting compound proposition;

2. A mechanism for generating new propositions from (i) knowledge already known and (ii) general statements called *rules*. This process is called *deduction*.

Symbolically, a proposition which is believed to be true may be represented as p. The belief that the same proposition is not true (false) is represented as $\sim p$. The opposite of a proposition being false is that the proposition is true, or $\sim (\sim p) \equiv p$. A situation in which p and $\sim p$ are claimed to be true is a *contradiction*. In classical logic, contradictions indicate an error in the logical system in which they occur.[1]

Propositions may be combined via conjunction (logical AND) or disjunction (logical OR). The symbolic representation of the conjunction of two propositions is:

$$p \wedge q \tag{1}$$

and that of the disjunction[2] is:

$$p \vee q \tag{2}$$

The laws of commutation and distribution for logical combination operators are respectively:

$$(p \wedge q) \equiv (q \wedge p) \tag{3}$$
$$(p \vee q) \equiv (q \vee p) \tag{4}$$

and

$$(p \wedge (q \wedge r)) \equiv (p \wedge q) \wedge (p \wedge r) \tag{5}$$
$$(p \wedge (q \vee r)) \equiv (p \wedge q) \vee (p \wedge r) \tag{6}$$
$$(p \vee (q \vee r)) \equiv (p \vee q) \vee (p \vee r) \tag{7}$$
$$(p \vee (q \wedge r)) \equiv (p \vee q) \wedge (p \vee r) \tag{8}$$

[1] Generally caused by an erroneous *rule*. Read on.

[2] The use of the symbol $\vee$ for disjunction is from the Latin word *vel*, meaning "inclusive OR". Unlike English, Latin possesses a separate word for "exclusive OR", *aut*.

Any combination of two propositions may be treated as a single proposition.

Negation applied to combinations leads to the following relationships, known as DeMorgan's Theorems

$$\sim(p \wedge q) \equiv (\sim p \vee \sim q) \tag{9}$$
$$\sim(p \vee q) \equiv (\sim p \wedge \sim q) \tag{10}$$

2.1 Rules

Deductive inference is accomplished through the use of *rules*.

> "A *Rule* is a hypothetical proposition composed of an antecedent and consequent by means of a conditional connective or one expressing reason which signifies that if they, viz. the antecedent and consequent are formed simultaneously, it is impossible that the antecedent be true and the consequent false."[3]

Translated into more modern terms, a rule is a conditional statement, consisting of an antecedent (the set of conditions to be met) and a consequent (the set of inferrences to be implied if the conditions required in the antecedent are satisfied).

While the laws of logic, such as the commutative law above, and the theorems derivable from them, such as DeMorgan's Theorems, are *domain independent*, rules are based on semantic information. Thus,

"If it is raining, then the runways are wet." (11)

is a rule. Such a rule is written formally as

$$p \supset q \tag{12}$$

where p represents the antecedent ("it is raining") and q represents the consequent ("the runways are wet"). The hypothetical nature of the statement is embodied in the symbol $\supset$.

Generally, rules have more than one antecedent, and may have more than one consequent. Rules may combine antecedents purely by conjunction, purely by disjunction, or in combination. As an alternative, a disjunctive rule can be split into several simple or conjunctive rules. While this is less efficient from a notational point of view, it avoids disjunction altogether. This is valuable when creating a computer program to do logic, because the program need only perform conjunction.

There are six forms in which deductive rules of inference may appear. For the purposes of this discussion, the description of two will suffice.

[3]This definition is from a translation of the fourteenth century logician Pseudo-Scotus (John of Cornubia) which appears in [1].

Modus Ponens This form, properly known as modus ponendo ponens, or the method of affirmation leading to affirmation[4], comes to the conclusion of the consequent if the hypothetical antecedent is declared. For the rule above, the information that it is raining leads to the inference that the runways are, indeed, wet. Formally

$$\begin{aligned} & p \supset q \\ & p \\ & \therefore \; q \end{aligned} \tag{13}$$

Modus Tolens The correct name is modus tollendo tollens, or the method of denial leading to denial.[5] In this form, the negation of the antecedent is concluded if the hypothetical consequent is declared **not** to be true. The formal definition is

$$\begin{aligned} & p \supset q \\ & \sim q \\ & \therefore \; \sim p \end{aligned} \tag{14}$$

A *rule* may be thought of as a generalization that can be applied to a domain of specific situations. In effect, rules are the analogs of algebraic equations in logic. This generalization is accomplished, as in algebra, through the use of *variables*.

Note that the rule stated in (11) is not really correct. Obviously, if the rain is falling in Boise, there is little effect on the runways at LaGuardia. The rule as stated is not precise enough. A more precise statement of the intended meaning of rule (11) is

"If it is raining *at some airport*, then the runways *at that airport* are wet." (15)

If the airport is represented by the variable x, that it is raining at x by Rx, and that the runways are wet at x by Wx, then the formal statement of rule (15) corresponding to (12) is

$$Rx \supset Wx \tag{16}$$

Further information on topics such as quantification can be found in [2].

3 Defeasible Logics and Truth Maintenance

In a classic paper in 1979, Jon Doyle [3] made a first attempt to create a computer program employing an extended classical logic for use with dynamic domains. Doyle's *Truth Maintenance* is based on the idea that reasoned inferences are supported by evidence. The evidence supporting an inference is composed of the propositions that were

[4]From the Latin *ponere*, "to affirm".
[5]From the Latin *tollere*, "to deny".

used as the antecedents of the rule that resulted in the inference. If one or more of the evidenciary propositions changes, then the inferred consequent must be examined to verify that it is still true. If an inference is no longer supported by any evidence, then the inference must be denied.

More formally, Doyle restricted the application domain to those systems in which all the propositions which satisfy modus ponens *also* satisfy

$$\begin{aligned} & p \supset q \\ & \sim p \\ & \therefore \sim q \end{aligned} \tag{17}$$

Propositions which meet this criterion are said to be logically equivalent. An example of logical equivalence is the state of a switch and the state of the voltage on a line controlled by the switch. When the switch is on, the voltage on the line is on, and while the switch is off, the voltage is off. While there are many examples of natural phenomena which exhibit this kind of behavior, there are many more which do not.

In Doyle's version of truth maintenance, evidence is kept in a "support list". In essence, a support list associated with an inference contains those propositions and a reference to the rule that were used to conclude it. Consider two rules which conclude the same proposition

$$R_1 : p \wedge q \supset c \tag{18}$$

and

$$R_2 : a \wedge b \supset c \tag{19}$$

These two rules are equivalent to a disjunction which embodies both rules

$$R_3 : (p \wedge q) \vee (a \wedge b) \supset c \tag{20}$$

If c is declared true, then its support list will contain one or the other (or both) of the conjunctions in the antecedent. If, at some future time, one of the propositions in the support list changes to false, the conjunction it appears within is removed from the support list. If the support list is empty, then by (17) c must be denied.

This scheme works well for the restricted set of domains which satisfy (17). Unfortunately, most processes occurring in nature **do not** fall into this set, and truth maintenance cannot be used successfully to reason about their dynamics.[6] Further, there is no explicit mention of time in the truth maintenance mechanism, so the length of time that a proposition might be true cannot be easily specified.

[6] In classical logic, the form described in (17) is considered to be an error, and is known as "denying the premise".

4 Default Logic

Classical logic treats knowledge in a very restricted way. There is a tacit assumption that the reasoner knows all that is necessary in order to proceed. While this may be true in mathematical proof, it is certainly *not* true in general. Human reasoners are often faced with *lack of knowledge*.

There is an ambiguity in the standard meaning of $\sim p$. This can be easily demonstrated. Suppose that p represents the statement, "It is raining". One interpretation of $\sim p$ is the statement, "It is *not* raining". Another interpretation is, "No information is available as to whether it is raining or not". In *either* case, it cannot be said that p is TRUE. For the purpose of mathematical proof, *either* meaning will suffice.

The domain of human reason is not as restricted as that of mathematical logic. The difference between knowing something is FALSE or TRUE and **not knowing** may be crucial. It is very important that the explicit meaning of $\sim p$ be well defined and understood.

Actually, there is implicit to the statement that some proposition is true the further statement that it is *known* to be true. When it is stated that, "It is raining", the actual *meaning* is that, "It is *guaranteed* that it is raining". Similarly, when some proposition is reported false, the meaning is that it is *certainly* NOT true.

Let us look more closely at the meaning of negation. There are statements such as, "Day is *not* night", in which the negation is a property of the domain. That is, in the case that the meaning of negation is the opposite state of the proposition, the veracity of the logical connection is semantically derived.

Alternatively, denial of possession of knowledge concerning the truth of a proposition has nothing to do with the domain of discourse. It is a purely *logical* matter, having only to do with form. Having or not having knowledge about a proposition has nothing to do with the semantic content of the proposition.

The importance of this distinction may be exhibited by reference to a rule used in the Modus Tolens progression (equation 14). If $\sim q$ means that q is *known* to be false, then the result of the progression is that p is *known* false as well. If, instead, $\sim q$ indicates that no knowledge is available about whether q is true (or false), then the result is that p is unknown, too. While the results look formally the same, the meaning is very different. In the case that the state of a proposition is known, we may not logically make any assumption concerning it. But when we have no direct knowledge about a proposition, it is often useful (or indeed, necessary) to make *assumptions*. For example, suppose that it is *known* that the temperature outdoors is 40°F, the dew point is just one degree less, and the humidity high. While it cannot be stated with certainty that there is fog, there is evidence to support the *assumption* that there is fog. Of course, if, in addition to the above, there is specific information that no fog is observable, an assumption about fog

should not be made.

This ability to make assumptions is the utility of supporting an "excluded middle". As a prerequisite, the two ambiguous meanings of **not** must be formally distinguished. For the remainder of this discussion, "$\sim$" will be used to denote that the opposite of the proposition following it is true, while "$\bar{\circ}$" will signify the lack of verifiable or trustworthy knowledge of the proposition which follows it.

A few rules concerning assumptions are necessary:

- Observation must always take precedence over assumption. This means that if an assumption has been inferred about some proposition, and contradictive information is subsequently observed concerning the same proposition, the assumption must be replaced by the observation.

- All propositions which are concluded from rules in which one or more of the antecedent propositions are assumptions, are themselves assumptions.

A new class of rules called **assumptive** rules may be defined. They are separated from normal rules because all their antecedents contain one (and only one) proposition claiming lack of knowledge about something, and they conclude an assumption. In the case of the "fog" example, let Fx indicate the proposition that there is fog at airport x. Let Qx indicate that the dew point is within two degrees of the outside air temperature at airport x. The assumptive (or default) rule may be written

$$(\bar{\circ}\, Fx) \wedge Qx \supset Fx^{\bullet} \tag{21}$$

to mean, "If it is not known that there is fog at an airport, but it is known that the dew point is close to the air temperature at that airport, it can be assumed that there is fog at that airport". The superscript asterisk appended to the conclusion is a reminder that this proposition is an assumption.

This particular choice of mechanism for making assumptions is a special case of Reiter's "Default Logic"[4].

The existence of assumptions extends the meaning of contradiction. To appreciate this, it is necessary to understand the different classes of information which can be present in a defeasible logic with default rules:

Observations are propositions that are obtained from outside of the logical system.

Inferred Facts are the consequent propositions resulting from rules in which all antecedent propositions are either observations or inferred facts.

Fundamental Assumptions are the consequent propositions arising from default rules.

Inferred Assumptions are consequent propositions which derive from rules in which one or more antecedent propositions are assumptions (of either variety).

The meaning of contradiction depends on the classes of the propositions involved in the inconsistency:

- Observation and observation: One of the observations must be in error. Given enough domain information (in the form of rules), it may be possible to ascertain which proposition to believe, but this is generally difficult even for humans.
- Observation and Inferred Fact: This situation almost always indicates an erroneous rule. Some rule in the deductive chain leading to the inferred fact must be responsible; with luck, there might only be one.
- Two Inferred Facts: A contradiction of this kind is also indicative of an error in a rule, but the reasoner will require outside assistance to determine the culprit, since there is no way to tell which rule is mistaken.
- Observation *or* Inferred Fact and Fundamental Assumption: The default rule that asserted the fundamental assumption is incorrect.
- Observation *or* Inferred Fact and Inferred Assumption: This is an interesting case. If only one of the propositions in the antecedent of the rule which inferred the assumption is an assumption, then the inferred assumption must be denied, and the assumption which appeared in the antecedent must be denied, and so on back to the fundamental assumption which started that chain of reasoning. This process is called "dependency directed backtracking". If it should happen that there were more than one assumption in the antecedent of any of the rules in the chain that led to the discrepancy, then a choice must be made: Which assumption should be retracted? There are various approaches that might be taken to answer this question:
  - One could retract the chronologically latest assumption and search for an alternative.
  - One could withdraw the chronologically earliest and search for an alternative.
  - Alternative assumptions could be found for each of the candidates, trying each one until a choice is found which does not cause the contradiction. This may be very time consuming, or, in fact, undecidable.
- Two Fundamental Assumptions: One of the default rules is incorrect.

In essence, all of these possibilities reduce to two major situations.

1. Contradictions among facts, which indicate errors of some kind.
2. Contradictions among assumptions, which require the replacement of one of the propositions with another assumption. The difficulty lies in deciding which assumption to replace, and with what.

The essence of the solution to the problem of replacing an assumption involved in a contradiction is the employment of a class of domain dependent preference rules. These rules have a general form of "If there is a contradiction involving two assumptions regarding X, then Y is the preferred assumption to retain (or to retract)."

The idea of preference rules may also be used to decide which of two (or more) conflicting observations to keep. If each observation is tagged with a description of its origin, then preference rules stating that one origin is more "believable" or more "important" may be used.

5 Temporally Dependent Propositions

Time dependence can be formally introduced into logic by defining

$$p\,(\tau) \tag{22}$$

to represent that proposition p is true during the time interval τ. The interval τ is a pair of numbers, such as Universal Times, such that the first member of the pair precedes or is equal to (i.e., is *before* or *at the same time of*) the second. Formally,

$$\tau = (t_1, t_2),\ t_1 \preceq t_2 \tag{23}$$

Having introduced this notion of the interval of veracity or the *activity interval* of a proposition, its effect upon all of the axioms of logic introduced in the previous section must be explored. It will suffice to examine only Conjunction, Disjunction, and the activity interval of the consequent of a rule.

The activity interval of a conjunct will be defined as the *intersection* of the activity intervals of the operands:

$$p\,(\tau_1) \wedge q\,(\tau_2) = p \wedge q\,(\tau_1 \cap \tau_2) \tag{24}$$

That this is a reasonable definition can be seen in the following example: If I am in room A during the time interval from two until four this afternoon ($p\,(\tau_1)$), and you are in room A during the interval from three until five this afternoon ($q\,(\tau_2)$), then *we* are in room A from three until four this afternoon ($p \wedge q\,(\tau_1 \cap \tau_2)$).

Nota Bene: This definition of conjunction effectively states that two propositions can interact if and only if they have overlapping time intervals. This may seem overly restrictive at first glance, especially considering the human penchant for describing many interacting events as following one another and being causally linked. However, closer examination reveals that the restriction is completely correct. Temporally disjoint events which *seem* to interact are invariably connected by some persistent process, produced as

an effect of the first event, which remains in effect at least until its time interval overlaps that of the second event.

The preceding motivates the definition of the activity interval of a disjunctive pair as the *union* of the activity intervals of the operands:

$$p(\tau_1) \vee q(\tau_2) = p \vee q(\tau_1 \cup \tau_2) \tag{25}$$

Again, using the room occupancy example: If I am in room A during the time interval from two until four this afternoon ($p(\tau_1)$), and you are in room A during the interval from three until five this afternoon ($q(\tau_2)$), then one or the other of us is in room A from two until five this afternoon ($p \vee q(\tau_1 \cup \tau_2)$).

5.1 Evanescence and Persistence

The rules for generating the activity interval of a logical combination of temporally constrained propositions allow the computation of the activity interval of the antecedent of a rule. However, the activity interval of the consequent of a rule is not necessarily the same as that of its antecedent. Processes and physical things whose activity intervals are shorter than the activity intervals of their antecedents are called *evanescent*. An example of an evanescent process is the firing of a gun. When the hammer drops the gun fires. The fact that the hammer remains down does not make the "bang" last longer. Other things and processes are persistent, lasting well after the events which created them have ceased to exist. For example, if the temperature is below freezing on the ground, and it is raining, ice will form on the ground. When the rain stops, the ice does not simply disappear.

The root of this problem is the domain dependence of the activity interval of a physical process. While it is true that no process can exist without some form of causal precedent, once formed a process may have an independent existence of its own. In the case of persistence, quite often the only way to "undo" something which has been "done" is to do something specifically designed to destroy it.

For practical purposes, there are only two classes of consequent

- Inferred propositions with activity intervals which are the same as the computed activity intervals of their antecedents. In this case, no further information is required in the consequent.

- Inferred propositions with independent activity intervals. Causality requires that the beginning of the consequent activity interval be the same as the start of the computed antecedent interval, but the domain dependent information to compute the end of the consequent activity interval must be supplied in the rule.

This topic will be discussed in greater detail in a forthcoming paper, "Implementation of the Tower Chief Planning System".

5.2 Resource Allocation and Planning

Classical temporal logic is insufficient to describe the domain and events which occur in the ATC environment characterized previously. In particular, the information available about the future state of some value may change during the execution of a plan. For example, there may be a weather prediction at 09:00 that claims that passage of a front with an associated shift in wind direction will occur between noon and one o'clock. A later prognostication, perhaps at 11:00, might change the time of the frontal passage or some parameter associated with it. Such a change may require a modification of a planned configuration shift which may already be in progress. Because of the infeasibility of certain transitions, the modification of one configuration choice may affect those which precede and follow it, and so on.

Preparation for the use of a specific configuration may demand the allocation of resources in advance. In winter, for example, one or more of the runways that are to be used in a future configuration may require snow or ice removal or treatment to prevent ice accumulation prior to being put into service.

In standard expositions on temporal logic, the processes that the system is designed to model usually involve the evolution of some physical quantity such as the position of a ball or the temperature of an object. The rules for this kind of modelling generally look like

$$p(t_1, t_2) \wedge q(t_3, t_4) \supset r(t_5, t_6) \tag{26}$$

where $p(t_1, t_2) \wedge q(t_3, t_4) \equiv p \wedge q\ [(t_1, t_2) \cap (t_3, t_4)]$. If r is a new state that was described previously by p, then $t_2 \preceq t_5$.[7] For example, "If some water is in the liquid state (p) during (t_1, t_2), and is brought in contact with a thermally massive object with a temperature greater than the boiling point of water (q) during (t_3, t_4), then the water will be in the vapor phase (r) during (t_5, t_6)."

A plan, on the other hand, is by definition something that is intended for future execution. Plans are based on what the planner *believes* is going to happen. Because the future is not fixed, a plan may have to be modified or even abandoned before or during its execution. Actions which are taken for the most part cannot be withdrawn when a plan is deserted. That such actions were based on beliefs which turned out to be false does not change the fact that they were executed. If the planner is to be able to explain the reasons behind its actions, it must recall its prior beliefs even if they were later proven to be wrong. In fact, the planner may change what it believes about the future as a result of its intention to carry out some plan.

Consider the reasoning involved in the allocation of some resource. Let $\mathcal{W}r(t_1, t_2)$ indicate that use of resource r is desired during the time interval (t_1, t_2). Similarly, let $\mathcal{A}r(t_1, t_2)$ indicate that resource r is available during the time interval, and finally let

[7] In other words, a thing cannot be in two disjoint states at the same time.

$\mathcal{H}r(t_1,t_2)$ represent that the resource has been allocated for the period. Then we might write

$$\mathcal{W}r(t_1,t_2) \wedge \mathcal{A}r(t_1,t_2) \supset \mathcal{H}r(t_1,t_2) \wedge \sim \mathcal{A}r(t_1,t_2) \tag{27}$$

to describe the rule for allocation: "If resource r is required during (t_1,t_2), and the resource is available during the period, then r is allocated for the interval, and is no longer available for allocation during that time." There is an apparent paradox in this formulation, since $\mathcal{A}r(t_1,t_2)$ and $\sim \mathcal{A}r(t_1,t_2)$ appear at the same time. The reason that this problem appears is that the reasoner's belief about the availability of the resource for other use during the specified interval changed as a consequence of the reasoning process itself.

The last example demonstrates that reasoning about the future may involve non-monotonic logic. However, propositions previously thought to be true cannot necessarily be simply retracted. They may be in the support lists of later propositions which involve actions or the reasoning process itself. A human reasoner in the resource allocation example would explain, "Of course, I believed the resource was available *before* I allocated its use. Now that I have committed the use of it during the time period, it is no longer available for other use." The reasoner is aware of the temporal order of the events of the reasoning process in addition to the projected time of execution of the plan.

There are *three* time periods associated with the resource allocation problem. First, the time in the future during which the use of the resource is desired. For the sake of easy reference, let us give that a name: the *activity interval*. Second, the time during which the planner believes that the resource will be available during the activity period. Third, the time during which the planner believes that the resource has been allocated for the activity period and will no longer be available for other use during that interval. Again, for reference, let's call these two periods "belief intervals". If we put the belief intervals into the logical statement of the resource allocation rule as subscripts to the propositions, we obtain

$$\mathcal{W}r_{(\tau_1,\tau_2)}(t_1,t_2) \wedge \mathcal{A}r_{(\tau_3,\tau_4)}(t_1,t_2) \supset \mathcal{H}r_{(\tau_5,\tau_6)}(t_1,t_2) \wedge \sim \mathcal{A}r_{(\tau_5,\tau_6)}(t_1,t_2) \tag{28}$$

and the paradox is resolved. The planner can now refer to what it believed prior to making its decision as well as its opinion after the act of making the decision.

Since a runway manager's duties include allocation of people and equipment to a variety of tasks, and a manager must take *actions* based on the current knowledge about the domain, the kind of reasoning described above is central to accomplishing the cognitive task of planning runway configurations.

6 Conclusions

The important attributes of this problem are not unique to runway configuration management. Temporal reasoning, default reasoning, and reasoning about the commitment

of resources are ubiquitous characteristics of almost all process management. Analysis of these kinds of cognition led us to a single representation and reasoning paradigm which integrates all three.

There is still much to be done. Currently, all input activity intervals must be clock times. It would be much more convenient to be able to enter times in qualitative terms by reference to information already known. Clock times often give a false sense of precision to information whose actual accuracy is fuzzy at best. It is well known that the process of maintaining a temporal database of the kind described here is $\mathcal{NP}$-hard.[8] What saves us is that the rate of change in the systems we have looked at is slow, and the airport effectively hardware resets every night. If this technology is to be employed in a broader spectrum of application, such as planning and scheduling of a planetary exploration robot, this efficiency problem will have to be solved.

7 Acknowledgements

I would like to thank the staff and administration of the Boston TRACON for their invaluable cooperation and assistance. Thanks also to the staff of the Flight Transportation Laboratory, especially Dennis F. X. Mathaisel and John D. Pararas, for many beneficial discussions.

[8] $\mathcal{NP}$-hard refers to the computational complexity of the problem. $\mathcal{NP}$ means "$\mathcal{N}$on-deterministic $\mathcal{P}$olynomial, indicating that the computation time for the brute force approach to the computation rises exponentially with the size of the problem. See [10].

References

[1] I. M. Bocheński. *A History of Formal Logic.* Chelsea Publishing Co., New York, 1970.

[2] Irving M. Copi. *Introduction to Logic.* MacMillan, New York, 1972.

[3] Jon Doyle. A Truth Maintenance System. *Artificial Intelligence*, 12:231–272, 1979.

[4] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[5] James F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, November 1983.

[6] James F. Allen and J. A. Kooman. Planning using a temporal world model. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 741–747, 1983.

[7] T. Dean and K. Kanazawa. Probabilistic causal reasoning. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, pages 73–80, 1988.

[8] Drew McDermott. A temporal logic for reasoning about processes and plans. *Cognitive Science*, 6:101–155, 1982.

[9] Yoav Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* MIT Press, 1988.

[10] G. Edward Barton, Robert C. Berwick, and Eric Sven Ristad. *Computational Complexity and Natural Language.* MIT Press, Cambridge, Massachusetts, 1988.

BUBBLES - An Automated Decision Support System for Final Approach Controllers

Zhizang Chi
Massachusetts Institute of Technology
Cambridge, MA

Assumptions-

1. An explicit Schedule exists for landings (and takeoffs) at each runway.
2. Each aircraft has declared an IAS for final approach and will be obligated to fly it as accurately as possible.
3. There is a continuous estimate of average windspeed on approach.

Objective-

PROVIDE AUTOMATED CUES TO ASSIST CONTROLLERS IN THE SPACING OF LANDING AIRCRAFT

The Cues have the following characteristics:

1. The cues are adaptive to estimation errors in position and speed by the radar tracking process, and piloting errors in execution of turns and commanded speed reductions.

2. The cues are responsive to the desires of the human controller;

   e.g., change landing sequence, insert a missed approach anywhere, insert planned take-offs between landings, increase/decrease spacing for any particular aircraft

The" merging area"

for aircraft arriving from different directions is on the runway centerline at 9-20 nm. from the runway. The schedule of landings and takeoffs is represented by a set of landing "bubbles" and takeoff "triangles" which move towards the runway at different approach speeds. Landing aircraft will intercept their bubble at different points on the extended runway centerline within the merging area.

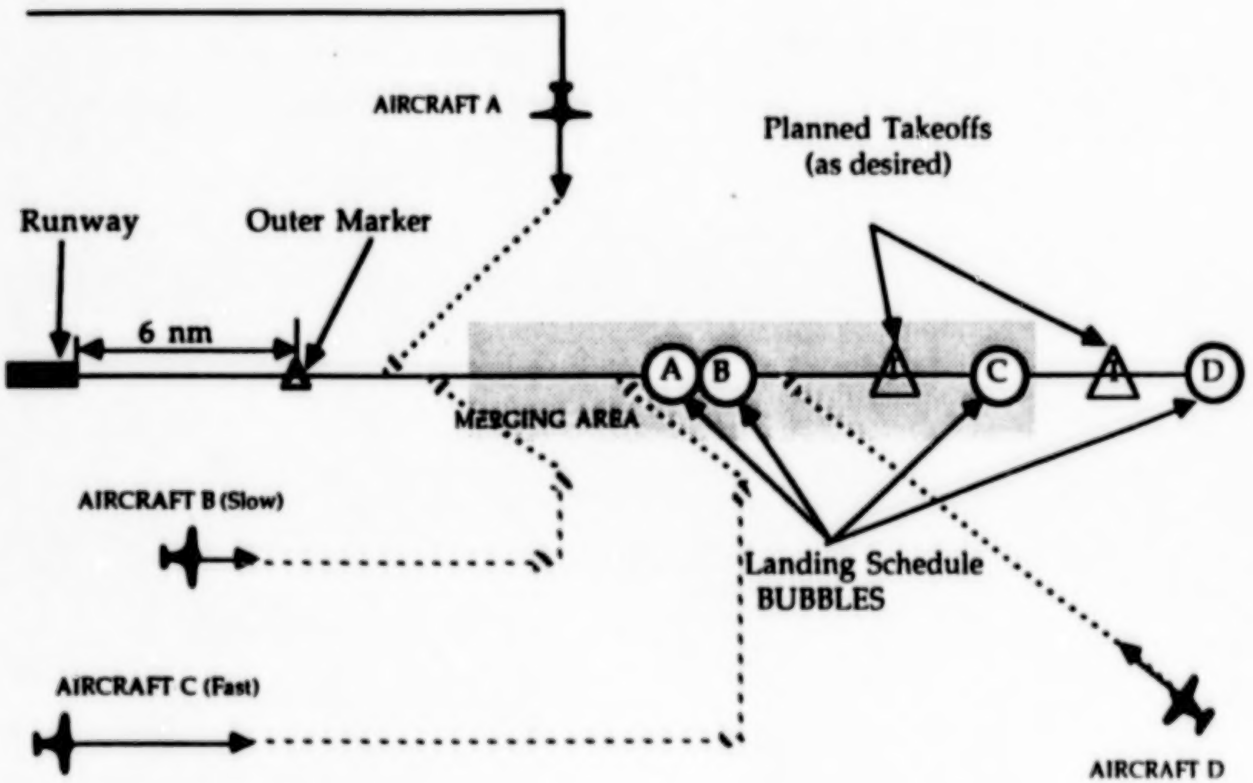


Figure 1- Region of Final Traffic Merging and Spacing

There are three planned spacing commands currently:

C1 - Turn to Base Leg

C2 - Turn to Intercept Leg

C3 - Reduce to Final Approach IAS

Cues are given for certain "spacing" commands for arrivals from different directions. At present, there are three such commands. A blinking cue gives a countdown to issue the command and anticipates the reaction time to execute.

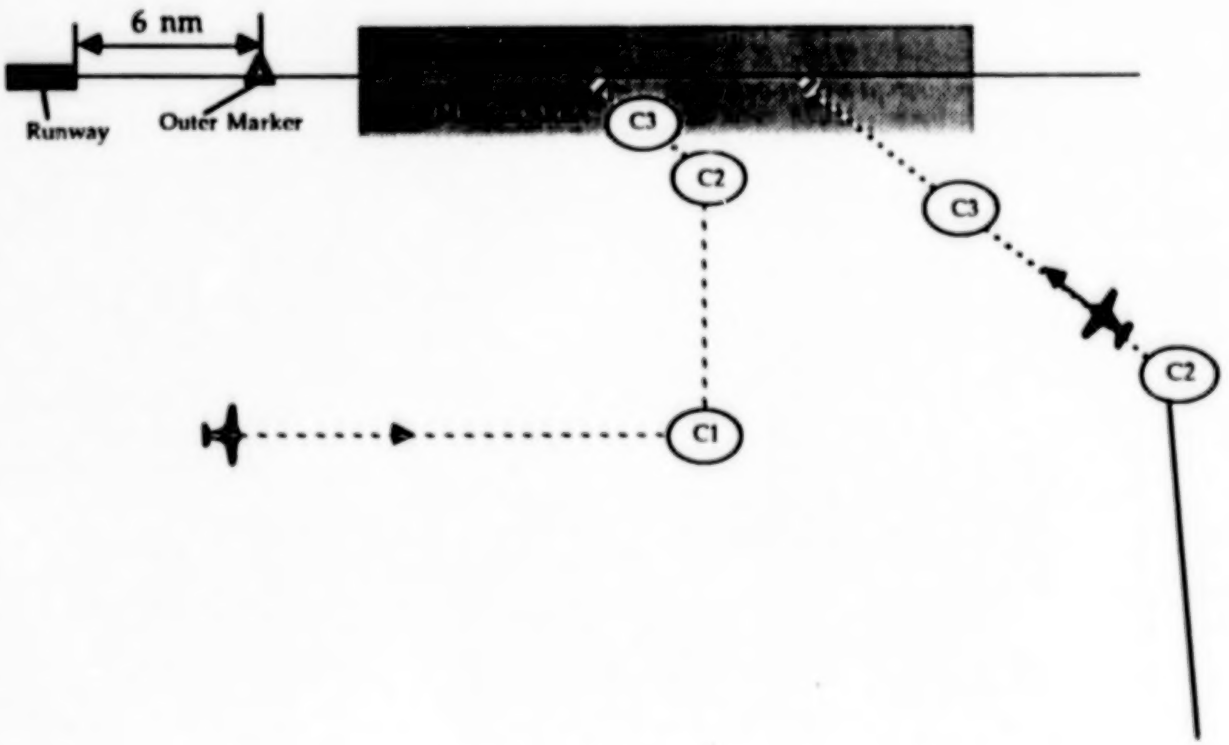


Figure 2 - Cueing of Spacing Commands

The correct Command is called whenever the aircraft reaches the α- wand. The angle α is constant given the final IAS for each aircraft. The aircraft can wander in speed and direction, but the correct cue is made whenever it reaches the wand. Wands will not be displayed to controllers.

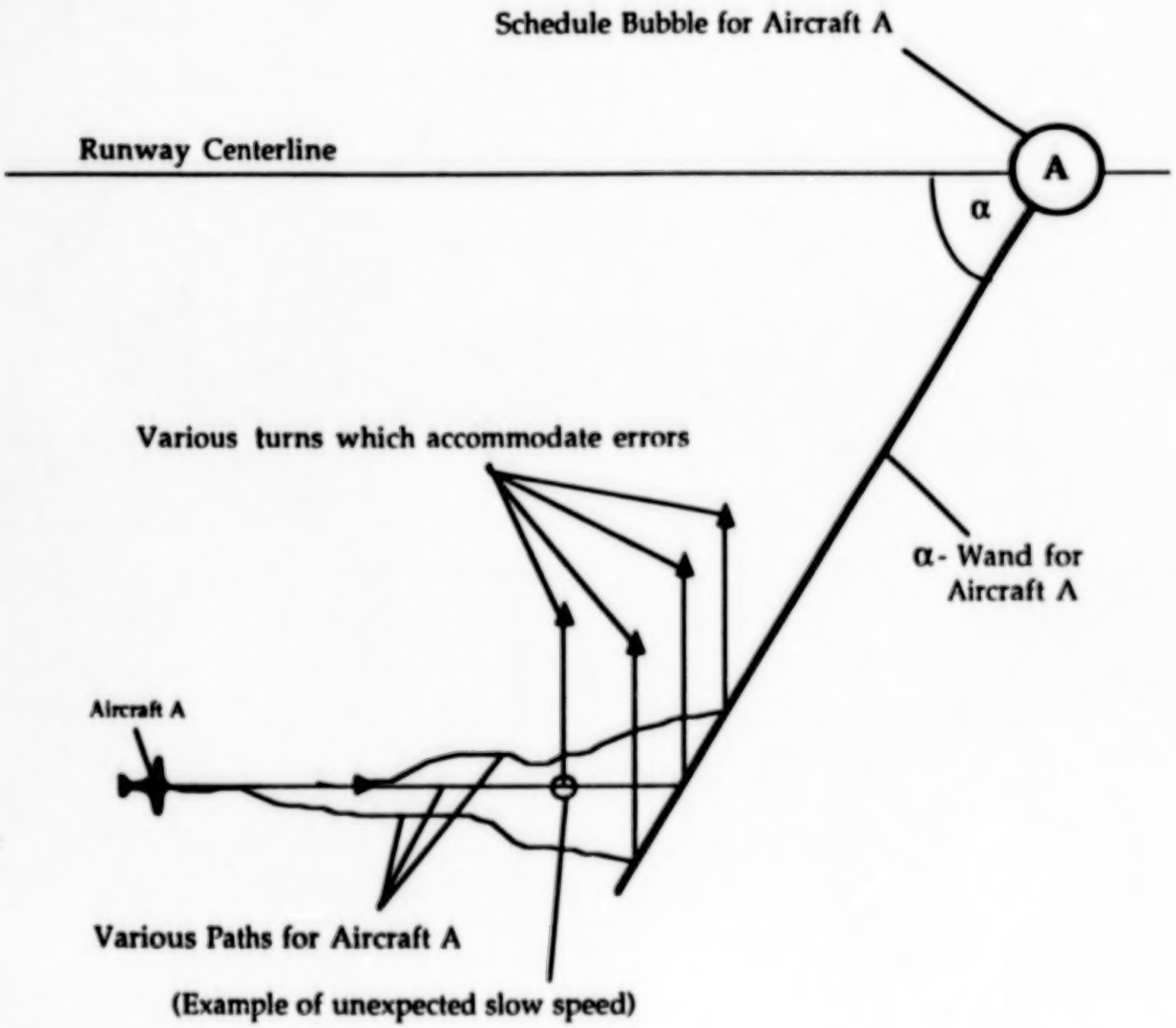


Figure 3 - Error Accommodation of the α-Wand

β remains a constant for correct spacing on the Base Leg. In this cue, the wand overtakes the aircraft at the correct time to turn to the intercept leg.

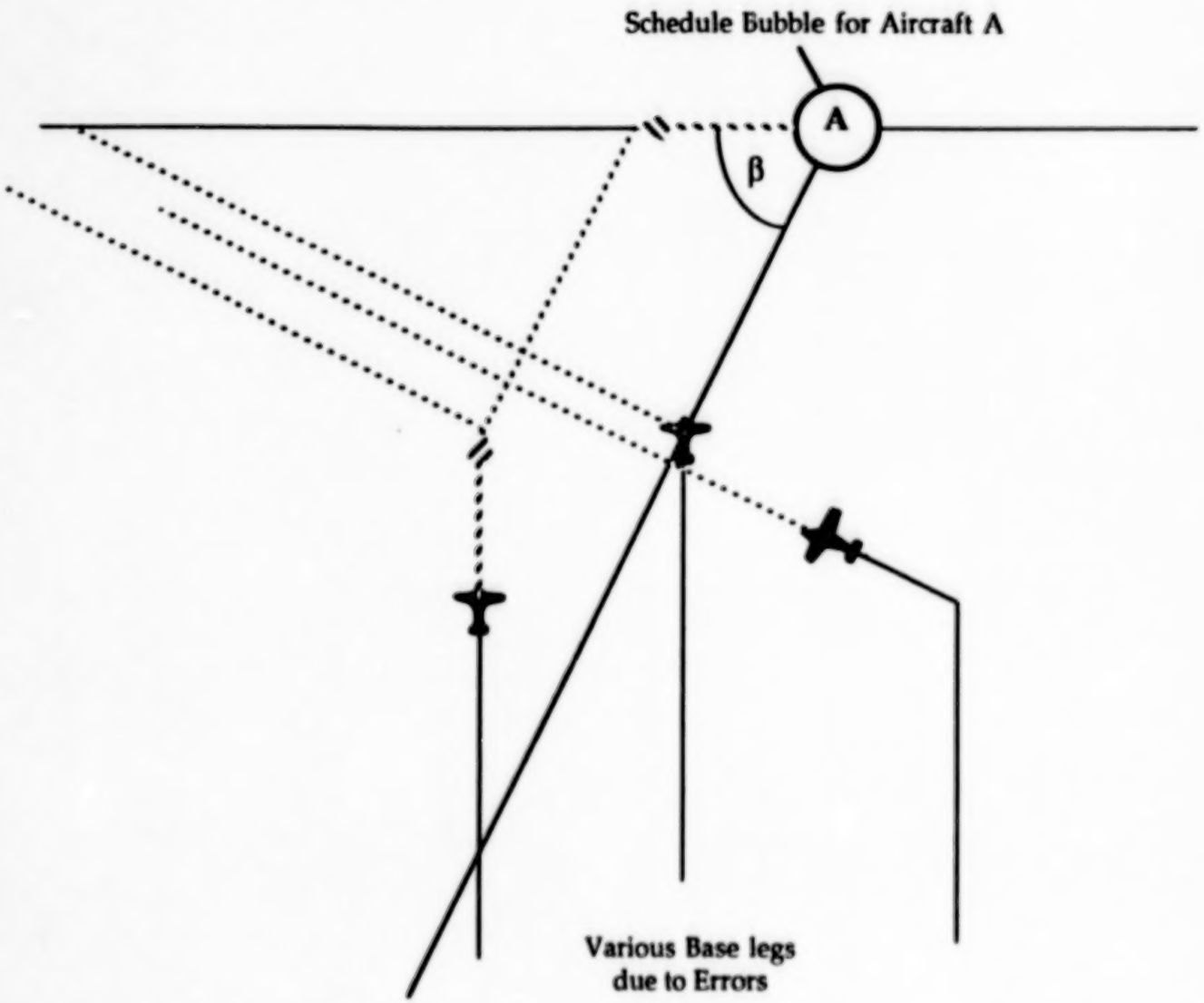


Figure 4 - Error Adaptation for the β - Wand

Speed Reduction called when aircraft catches up to the γ -wand on the intercept leg close to the runway centerline. Aircraft are kept above their declared final approach speed by 10-20 knots during earlier maneuvering. Wind speed on approach must be continuously updated.

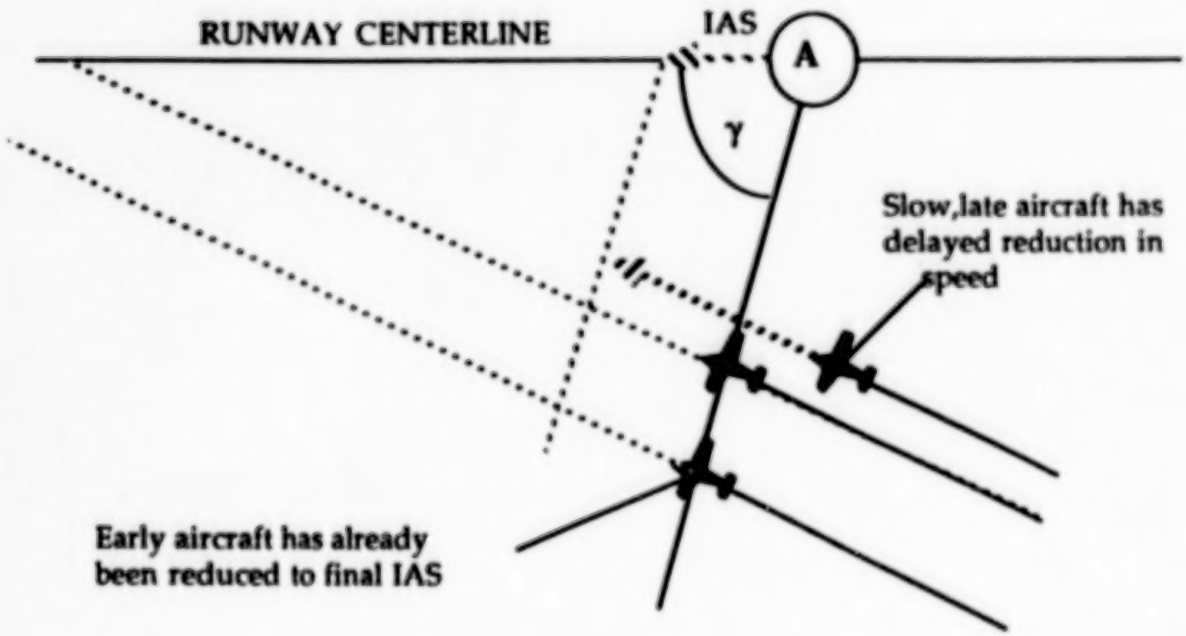


Figure 5 - Error Adaptation for the γ -Wand

Any Bubble can be selected and moved graphically to a new position by the Controller. The Spacing Cues will automatically adjust. The Bubble will blink or change color if the move is not feasible.

A complete slide of all subsequent aircraft can also be done. The planned insertion of waiting takeoff aircraft can be displayed during moves of landing aircraft. The planned landing schedule can be set to automatically insert another aircraft by a small opening of the landing spacing. This maximizes total operational rate of the runway.

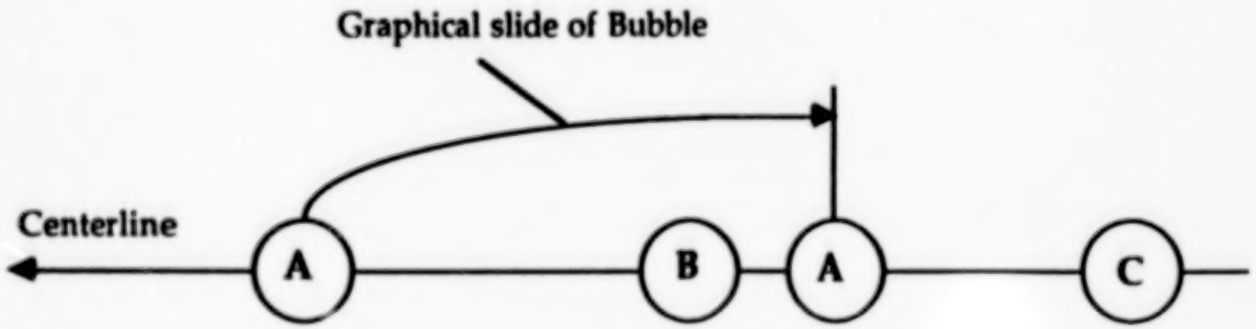


Figure 6 - Changing the Desired Sequence or Spacing

Currently, we are introducing a method of automatic adaptation to any residual errors between the aircraft and the bubble as the aircraft reaches the runway centerline.

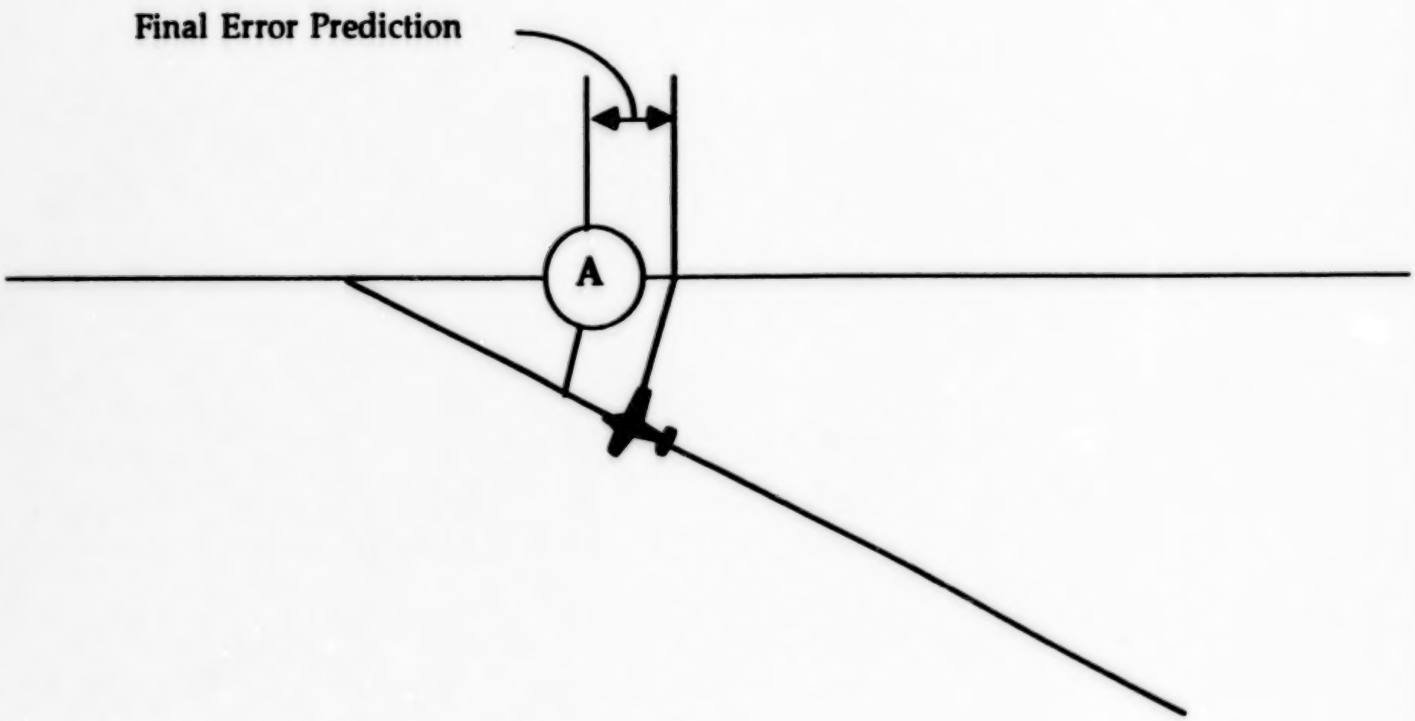


Figure 7 - Adaptation to Centerline Errors

Final Error causes an automatic shift of bubble to actual aircraft position if late and, if necessary, automatically shifts all subsequent bubbles within the limits of their feasible moves. This keeps the scheduled bubble positions tied to actual performance of the aircraft over a longer period.

ANALYSIS OF AIRCRAFT PERFORMANCE DURING LATERAL MANEUVERING FOR MICROBURST AVOIDANCE

Denise Ávila de Melo*
Embraer S.A. - Brazilian Aeronautical Enterprise
São José dos Campos, São Paulo BRAZIL

and

R. John Hansman, Jr. †
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, MA

Abstract

Much of the prior research on aircraft escape procedures during microburst encounters has assumed that the aircraft penetrates the center of the microburst in straight flight. The microburst core is the region where the strongest head to tail wind and downburst are present. Aircraft response to a severe and a moderate 3-dimensional microburst model using nonlinear numerical simulations of a Boeing 737-100 was studied and the relative performance loss was compared for microburst escape procedures with and without lateral maneuvering. The results showed that the hazards caused by the penetration of a microburst in the landing phase were attenuated if lateral escape maneuvers were applied in order to turn the aircraft away from the microburst core rather than flying straight through. If the lateral escape maneuver was initiated close to the microburst core, high bank angles tended to deteriorate aircraft performance. Therefore, lateral maneuvering should be employed if the position of the microburst is know, however, only low bank angles should be applied once the core has been penetrated. Lateral maneuvering was also found to reduce the advanced warning required to escape from microburst hazards but required that information of the existence and location of the microburst is available (ie., remote detection) in order to avoid an incorrect turn toward the microburst core.

1. Introduction

Low-altitude wind shear presents a significant hazard to aircraft during landing and take-off operations. Severe microbursts, storm downdrafts, which are small in horizontal cross sections and highly transient, present the greatest danger to aircraft, ranging from small general aviation aircraft to jet transports. The risks posed by all forms of wind shear can be reduced if information is available to warn the pilot about the presence of low level wind shear and if the pilot has the best available information on escape techniques. Most of the prior research on aircraft flight dynamics and microburst escape procedures has focused on longitudinal dynamics. This is equivalent to the aircraft penetrating the center of the microburst where only the effect of the horizontal head to tail wind components and the vertical downburst are considered. Currently, the FAA Windshear Training Aid[1] recommends microburst escape procedures which are limited to maneuvers in the longitudinal plane. With the advent of systems which can remotely detect microburst such as Doppler weather radar, the possibility of lateral maneuvering for microburst avoidance should be considered.

Based on the assumption that information about existence and location of a microburst is available, studies were conducted to evaluate the relative performance loss and recovery capability for microburst escape procedures with and without lateral maneuvering. Severe and moderate microbursts cases were considered. From the simulation results, recommendations were made for improving microburst recovery capability.

2. Method of Approach

2.1 Equations of Motion

The set of nonlinear equations of motion which describe the aircraft dynamics in the 3-dimensional space are derived in Ávila de Melo[2] in the inertial velocity axes following Psiaki and Stengel's[3] procedure and using the notation of Etkin[4].

2.2 Aircraft Data

The simulations used a simplified nonlinear aerodynamic model of the Boeing 737-100 (NASA Langley ATOPS research aircraft)[5]. The power plant dynamics were approximated as a first order model with a time response of 2 seconds up to the maximum thrust of 13,000 lb.

$$\dot{T} = (\delta T - T)/T_R; \; T_R = 2 \text{ sec.}$$

In the initial condition, the aircraft was assumed to be at a constant airspeed of 130 kts, on a 3°glide slope, with angle of attack of 1.3°and weight of 80,000 lb. Landing gear and flaps are in the landing configuration. The trim positions of the control devices were the following:

δ_e = 2.9° - elevator deflection;
δ_a = -5° - spoiler deflection;
δ_a = 0° - aileron deflection;
δ_r = 0° - rudder deflection;
T = 8,081 lb - total thrust.

2.3 Microburst Model

The microburst model used for this study is similar to the 3-dimensional microburst model by Oseguera and Bowles[6]. However, for simplicity, the model was made invariant with altitude. The maximum intensity horizontal and vertical velocity profiles were used to represent a worst case. It should be noted that this somewhat exaggerates the hazard since the maximum horizontal and vertical intensity do not normally occur at the same altitude. Details on the analytic microburst equations are given in Avila de Melo[2].

*Engineer, Flight Mechanics Group
†Associate Professor, Associate Fellow AIAA

Copyright © 1989 by Denise Ávila de Melo and MIT. Published by the American Institute of Aeronautics and Astronautics, Inc. with permission.

Preprint AIAA 28th Aerospace Sciences Meeting
January 8-11, 1990 AIAA-90-0568

2. 3. 1 Microburst Types

Two different microburst magnitudes were modeled by specifying the radius of peak outflow and the maximum wind velocity. The radius of the downdraft is assumed in the Oseguera and Bowles model[6] to be approximately 89% of the radius of peak outflow. In this work the microburst core is considered to be the region from the peak head wind to the peak tail wind, ie., within the peak outflow velocity contour.

• Severe Microburst: The severe case was based on the Andrews AFB event (Camp Spring, Maryland on August 1, 1983)[7] with a wind shear intensity approximately 120 kts in 4,000 ft. The maximum horizontal velocity was assumed 60 kts and the maximum vertical velocity was 45 kts (see Fig. 2-1).

• Moderate Microburst: The moderate case was based on the microburst encountered by Delta Airlines Flight 191 (Dallas/Ft. Worth Airport on August 2, 1985)[7] with a wind shear intensity of approximately 60 kts in 4,200 ft. The maximum horizontal velocity was 30 kts and the maximum vertical velocity was 22 kts.

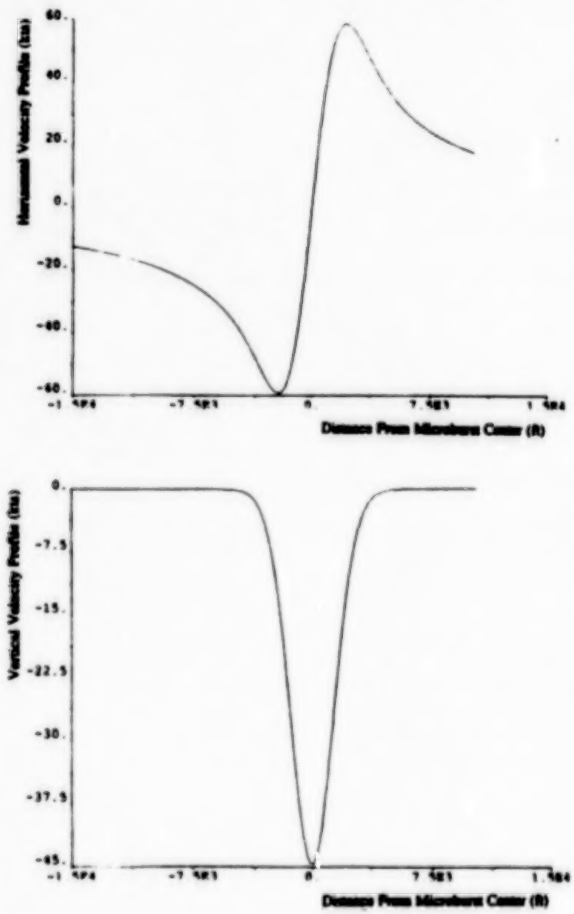


Fig. 2-1 Horizontal and vertical velocity profiles for the severe microburst case.

2. 4 Simulation Description

2. 4. 1 Control Laws And Guidance Schemes

Linear automatic control laws as described by McRuer *et al.* [8] were used in order to simulate pilot control actions or to stabilize the nonlinear aircraft model. For the pitch altitude control, integral pitch angle feedback was used to control elevator deflection with integrator. The roll attitude control was accomplished by feeding back roll attitude and roll rate to control aileron deflection.

The longitudinal-only escape maneuver was based on the FAA Windshear Training Aid[1] . It consisted of:

- Rotating the aircraft to 15° of pitch angle.
- Applying maximum thrust.
- Maintaining landing gear and flap configuration.
- Respecting stick shaker (ie., not allowing angles of attack greater than the stall angle).

In the longitudinal-only escape maneuver a maximum thrust equal to 13,000 lb per engine was applied, the stabilizer was deflected to -3.6° and the elevator was controlled to maintain either a pitch angle of 15° or the incipient stick shaker. In the cases where the aircraft flew off the microburst axis of symmetry, automatic lateral control was required in order to keep the aircraft flying straight and avoid the aircraft rolling towards the microburst due to the strong cross-wind component.

In the lateral escape maneuver, in addition to the longitudinal procedure described above, a step command in aileron was initially applied within an approximate roll rate of 5°/sec. When the desired bank angle was reached, the automatic lateral control was used in order to maintain the desired bank angle. When the aircraft had turned 90° in heading, opposite aileron was applied so that the bank angle reduced to zero. Sufficient rudder was applied to keep sideslip angle close to zero in the case when there was no microburst (ie., normal coordinated turn).

2. 4. 3 Nonlinear Model Description

Nonlinearities such as aerodynamic stall characteristics play important roles in limiting microburst penetration capabilities. Therefore, the simulation of the aircraft nonlinear equations of motion and aerodynamic data was essential. Also, deviations from steady motion were not small and the longitudinal and lateral motions were coupled. It was, therefore, necessary to use the complete set of nonlinear equations of motion in the simulation.

A program containing the aircraft nonlinear equations of motion, the microburst analytic equations and the aircraft aerodynamic data was used to simulate the aircraft flight through the 3-dimensional microburst. In the program, the nonlinear differential equations were solved by a hamming-predictor corrector integration routine[9]. The control laws and guidance schemes were incorporated in the program in order to simulate the escape maneuver actions a pilot would take and to stabilize the aircraft or keep it flying at the desired attitude. The output of the program is in the form of plots of the state variables.

3. Flight Through The Microburst Axis of Symmetry

In each set of simulations, the performance of the aircraft employing lateral maneuvering to avoid the microburst was compared with the performance of the aircraft flying straight through the microburst and applying only the FAA Windshear Training Aid[1] recommended escape maneuver. Four bank angles were used in the lateral maneuvering cases: 5°, 10°, 15° and 20°. The aircraft was initially flying in a trajectory along the microburst axis of symmetry which penetrated the microburst core (see Fig. 3-1). The escape procedure was initiated at various distances away from the microburst core.

3. 1 Severe Microburst Case

In the severe microburst case, five sets of simulations were run. In the first set, the escape maneuver was initiated at the distance of 10,000 ft from the microburst center. This was the point where, due to the outflow, an increase in airspeed of approximately 15 kts was detected. This is one of the recommended microburst recognition criteria suggested by the FAA Windshear Training Aid[1]. In the last set, the escape procedure initiated at the microburst center where the aircraft experienced the greatest loss of altitude and change in vertical

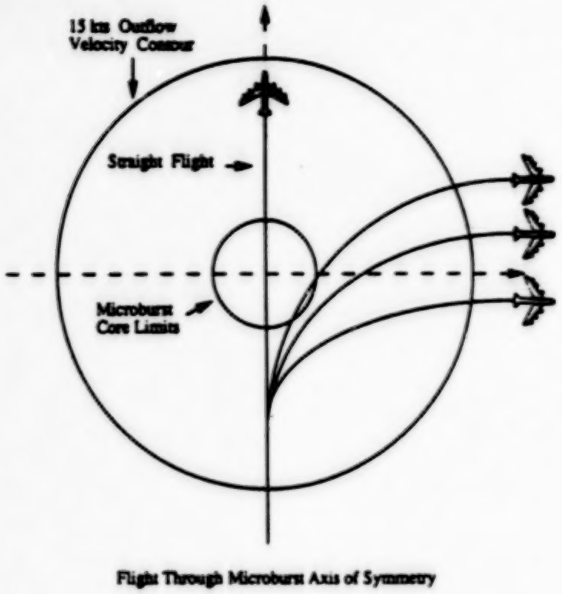


Fig. 3-1 Flight through the microburst axis of symmetry.

speed. In between these two extreme cases, three other intermediate points of escape maneuver initiation were considered. The point at 2,000 ft from microburst center was approximately the point of peak outflow velocity and peak airspeed increase. The points at 6,000 ft and at 4,000 ft from microburst center were studied in order to check the effects of initiating the escape maneuver while the aircraft was transiting the region of increasing head wind and before it got into the microburst core.

An example is shown in Fig. 3-2 and 3-3 which plot of the aircraft horizontal trajectory and altitude versus time, respectively, with and without lateral maneuvering for the case when the escape procedure was initiated at the point of peak outflow velocity. This point was approximately 2,000 ft from the microburst center. In all escape maneuvers, the aircraft fell below the 3° glide slope line and was in stick shaker condition from the point of escape initiation. In the 20° bank angle maneuver, the aircraft loss of altitude was greater than in the straight flight. The major reason for that was the very intense tail wind (Fig. 3-4) combined with the loss of lift due to steep bank angle. There was not a significant improvement in performance for the 5° of bank case since the aircraft trajectory continues to cross most of the region of high head to tail wind shear and intense downburst. Some improvement in performance occurred with 10° of bank angle and the best recovery capability was achieved by banking the aircraft to 15°. The tendency of the aircraft to weather cock towards the microburst center for low bank angles can be seen in the trajectory plot of Fig. 3-2.

Recovery is defined as the point where the aircraft assumes a steady state positive climb rate.

Fig 3-5 shows a comparison of the aircraft altitude versus time plots when the escape maneuvers were initiated at 6,000 ft, 4,000 ft, 2,000 ft (peak outflow velocity) from the microburst center and at the microburst center. For escape procedure initiated before the peak outflow velocity point, lateral maneuvering presents better recovery capability than the straight flight. Lateral maneuvers take the aircraft away from the microburst core where there are the strongest head to tail wind and downburst.

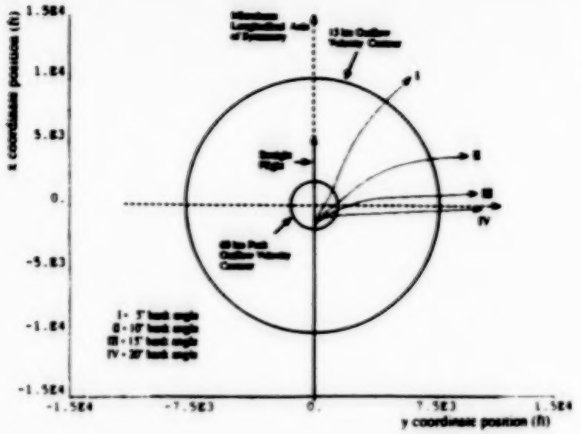


Fig. 3-2 Aircraft trajectory for escape maneuver initiated at peak outflow velocity of severe microburst.

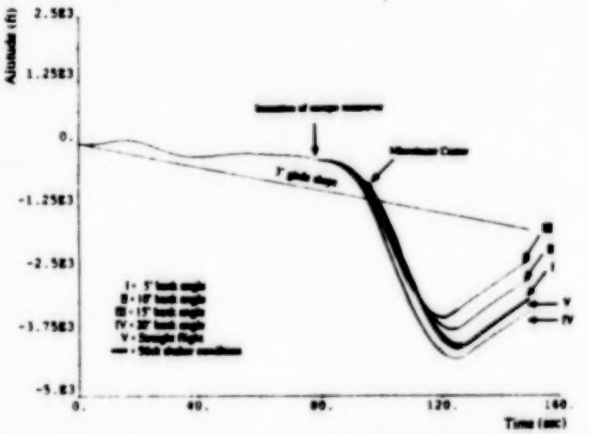


Fig. 3-3 Aircraft altitude for escape maneuver initiated at peak outflow velocity of severe microburst .

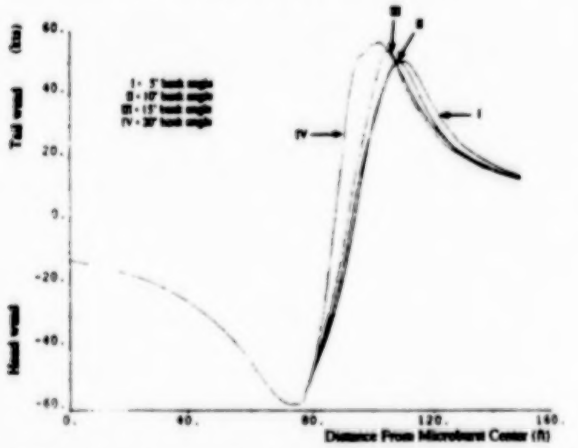


Fig. 3-4 Head to tail wind velocity profiles experienced by the aircraft in the various bank angle trajectories for escape maneuver initiated at peak outflow velocity of severe microburst.

Figs. 3-5 and 3-6 show that the aircraft performance varies with the bank angle but for the cases where escape maneuver was initiated before and at the peak outflow velocity point, banking the aircraft to 10° showed enough capability to fly away from the microburst core and significant performance improvement when compared with the straight flight. Banking the aircraft to smaller angles did not show a significant improvement in recovery capability, while larger bank angles tend to deteriorate aircraft performance. The deeper the aircraft is in the microburst core, the greater is the performance degradation due to performance loss at steep bank angles.

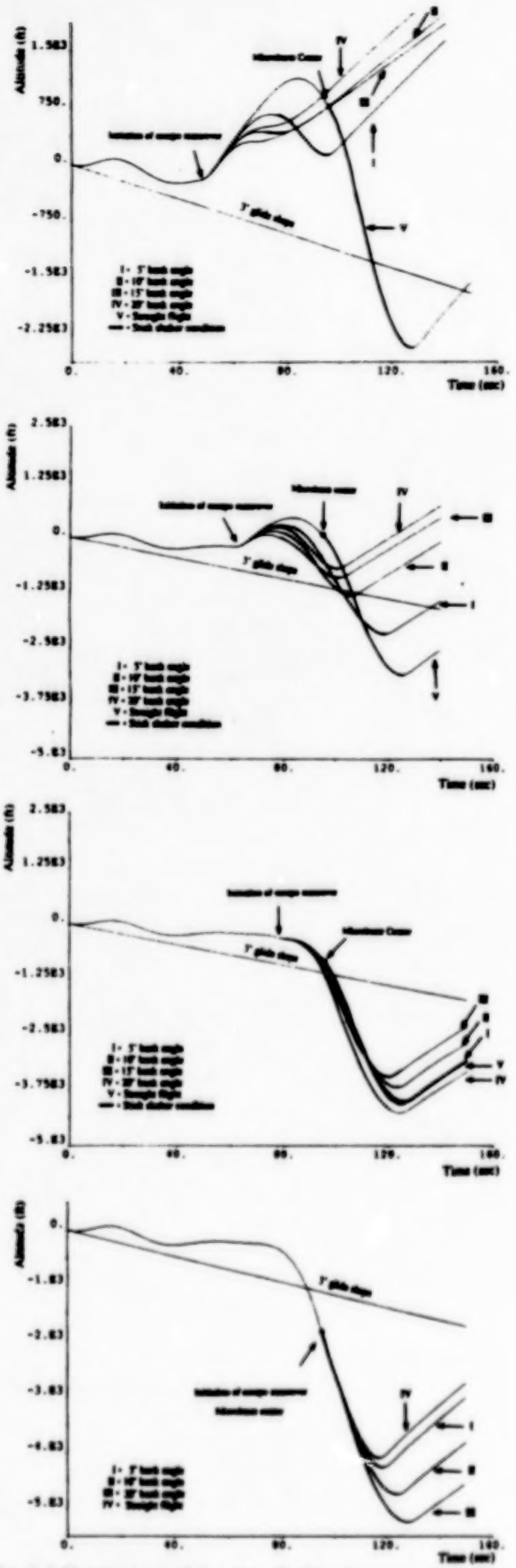


Fig. 3-5 Comparison of the aircraft altitude versus time plots when escape maneuver is initiated at a) 6,000 ft , b) 4,000 ft, c) 2,000 ft (peak outflow velocity) from microburst center and d) at microburst center.

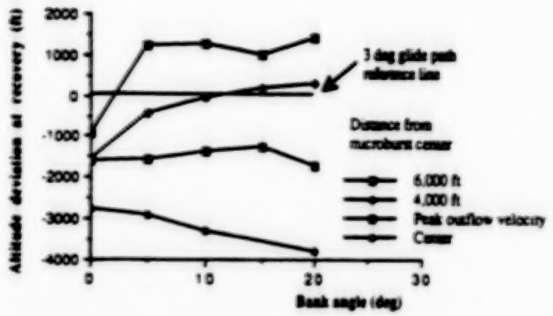


Fig. 3-6 Aircraft altitude deviation from the 3° glide slope path at the point of recovery. Positive values mean aircraft is above 3° glide slope line, negative value means below.

3. 2 Moderate Microburst Case

In the moderate microburst, five sets of simulations were run. In the first set, the escape maneuver was initiated at the distance of 6,000 ft from the microburst center, where an increase in airspeed of approximately 15 kts was detected. This is one of the recommended recognition criteria suggested by the FAA Windshear Training Aid[1]. In the last set, the escape procedure was initiated at the microburst center, the point where the aircraft experienced the greatest altitude loss and change in vertical speed. In between these two extreme conditions, three other intermediate points of escape maneuver initiation were considered. The point at 2,360 ft from the microburst center was approximately the point of peak outflow velocity and peak airspeed increase. The points at 4,000 ft and 3,000 ft from microburst center were studied in order to check the effects of initiating the escape maneuver while the aircraft was transiting the region of increasing head wind and before it got into the microburst core. Since it is a moderate microburst, the effects of the wind components in the aircraft performance were not as strong as in the severe case discussed in section 3. 1.

Fig. 3-7 shows the behavior in the moderate case similar to the observed for the severe case with significantly less performance loss.

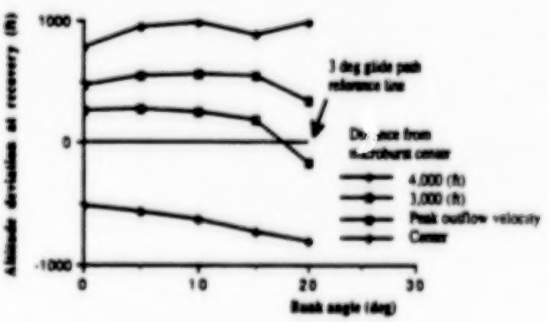


Fig. 3-7 Aircraft altitude deviation from the 3° glide slope path at the recovery. Positive values mean aircraft is above 3° glide slope line, negative value means below.

3. 3 Analysis and General Results

For the cases studied , a bank angle of 10° appeared to provide the best average performance. Banking too steeply causes performance loss particularly if the aircraft is deep in microburst core. The deeper the aircraft is in the microburst the less it should be banked and if it has penetrated the core, no lateral maneuver should be applied.

Lateral maneuvering reduced the advanced warning required to escape from the microburst . An example is shown in Fig. 3-8 for the severe microburst. The straight flight escape maneuver must be initiated at 10,000 ft from the severe microburst center to provide approximately the same recovery capability as 10° bank lateral maneuver initiated at 4,000 ft from the severe microburst center. This 6,000 ft difference represents 38 sec less required warning time in the simulation.

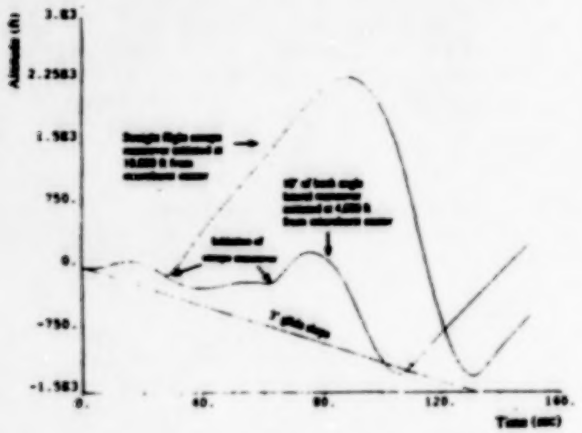


Fig. 3-8 Comparison of aircraft altitude for straight flight escape maneuver initiated at 10,000 ft from severe microburst center and for 10° of bank angle when t the lateral maneuver was initiated at 4,000 ft from the severe microburst core.

4. Flight Off The Microburst Axis of Symmetry

Simulations were run in order to evaluate the hazards associated with an incorrect lateral maneuver which would take the aircraft towards the microburst core, as shown in Fig 4-1. In each set of simulations, the aircraft was assumed to initially penetrate the microburst on a trajectory offset from the microburst core. Three situations were considered: 1) The aircraft turned towards the microburst core, 2) The aircraft turned away from the microburst core, 3) The aircraft flew straight ahead in an offset trajectory parallel to the microburst axis of symmetry.

The simulations were run considering only the points of escape maneuver initiation and the bank angle values that in the case of a turn towards the microburst core would take the aircraft in a trajectory that crossed the microburst core. This was an approximation of the worst case.

During the simulations, it was observed that direct axisymmetric penetrations of the microburst core have been found to be less hazardous than those which were slightly offset from the core. The difference in performance was not significant and was thought to be a result of cross wind component effects. Also, it was necessary to use automatic control to maintain the roll angle close to zero and to compensate for the rolling effects of the wind, which were quite strong in the severe microburst case. No calculations were done in order to evaluate stick forces, which with the lateral maneuvering may be very high in order to counter act natural airplane pitching tendencies due to the cross wind and to airspeed and lift loss.

4.1 Severe Microburst Case

Two sets of simulation were run. In the first set, the escape maneuver was initiated at a lateral position of 430 ft from the microburst axis of symmetry and at 2,000 ft from microburst center. In the second set, the escape maneuver was initiated at a lateral position of 3,000 ft from the microburst axis of symmetry and at 4,000 ft from microburst center.

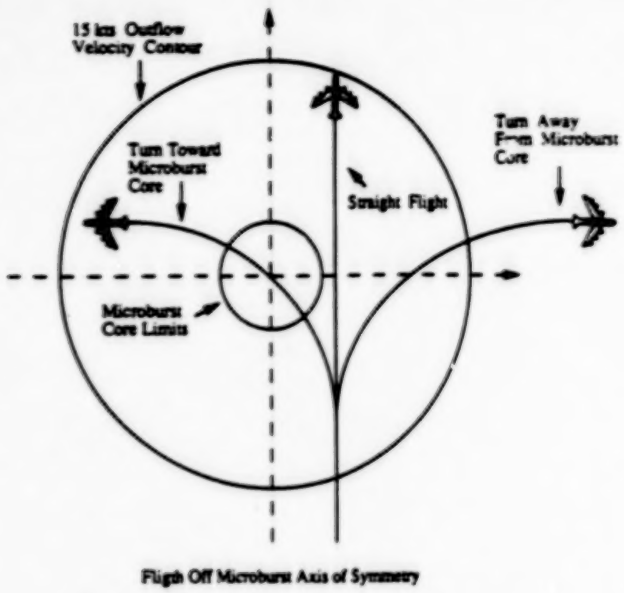

15 kts Outflow Velocity Contour
Turn Toward Microburst Core
Turn Away From Microburst Core
Straight Flight
Microburst Core Limits
Fligth Off Microburst Axis of Symmetry

Fig. 4-1 Flight off microburst axis of symmetry.

4.1.1 Flight at a Lateral Displacement of 430 ft

Fig. 4-2 shows a comparison of the aircraft altitude when, in a 10° bank angle, the aircraft turned away or towards the microburst center. In the case of the incorrect lateral maneuver, the aircraft trajectory took it into the microburst core. Consequently, the aircraft was taken to an altitude 40 ft lower than the straight flight trajectory and 750 ft lower than the turn away from the microburst core trajectory. There is a significant advantage in turning away from microburst core and relatively minor loss in performance due to the turn towards the core.

In this severe microburst case, the weather cocking effect was very strong due to the high intensity of the cross wind velocity at this lateral position. For this reason, it was impossible to simulate the aircraft flying in straight line through the microburst. Very high automatic control gains were necessary in order to avoid the aircraft's tendency to turn into the microburst center. The very high gains increased significantly the computational time. Therefore, the highest gain value which did not sacrifice too much the computational time was used to keep the aircraft flying as close as possible to straight flight. In real situation, the high required gains would represent a significant increase in pilot's workload in order to make course corrections required to keep the aircraft aligned with the run way centerline.

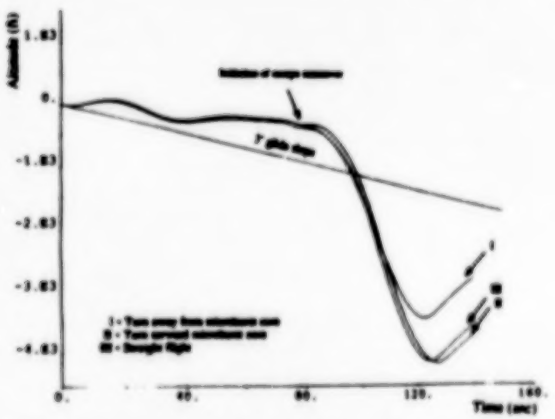


Fig. 4-2 Comparison of the aircraft altitude when the aircraft turns away or towards the severe microburst center and when it flies straight through the microburst - lateral displacement of 430 ft.

4.1.2 Flight at a Lateral Displacement of 3,000 ft

Fig. 4-3 shows a comparison of the aircraft altitude when, in a 15°of bank turn, the aircraft turned away or towards the microburst center. For this large initial displacement, the recovery capability in straight flight was about 900 ft larger than in case of an incorrect lateral maneuver towards the microburst core. However, it should be noted that this represented the worst case of incorrect lateral maneuver.

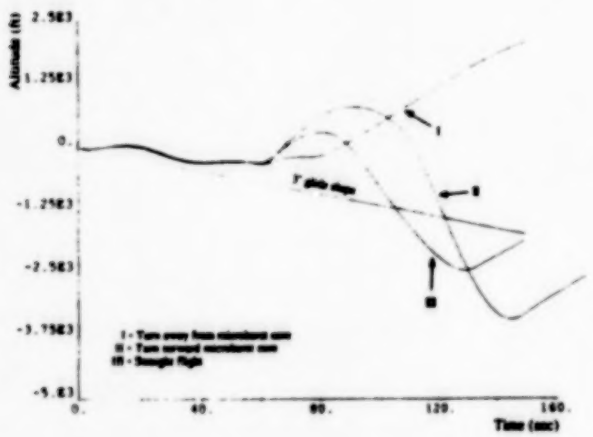


Fig. 4-3 Comparison of the aircraft altitude when the aircraft turns away or towards the severe microburst center and when it flies straight through the microburst lateral displacement of 3,000 ft.

In this severe microburst , the weather cocking effect of the cross wind was still strong even at this larger lateral distance. However, the very high control gains were not necessary to avoid the aircraft's tendency to turn into the microburst center.

4.3 Analysis And General Results

The moderate microburst case showed similar results except that the difference in performance loss between the turn towards the microburst center and the straight was more accentuated.

The recovery capability when the aircraft turned away from the microburst core was, in all cases, greater than when the aircraft flew straight or turned towards the microburst core. As expected, turning towards the microburst core represented a decrease in performance, since besides experiencing the strongest wind shear and downdraft, the aircraft suffered performance deterioration due to the bank angle. It shows that these incorrect trajectories were selected to penetrate the core and, therefore, reported the worst case of incorrect lateral maneuvering. Also, if the escape maneuver was initiated at larger lateral distances from the microburst center than the ones evaluated in these simulations, the risk associated with an incorrect turn towards the microburst core appeared to be relatively minor since the aircraft would have already have gained sufficient altitude before getting to the microburst core.

5. Conclusions

In summary, the simulations found:

- Significant improvement in the escape capability was obtained with moderate lateral maneuvering. For the aircraft studied in this work, the optimal bank angle for the cases studied appeared to be 10°. Banking the aircraft to smaller angles did not show a significant improvement in recovery capability, while larger bank angles tend to deteriorate aircraft performance.

- The deeper the aircraft penetrated the microburst core, the less bank should be applied. Significant performance improvement was obtained in comparison with the straight flight if lateral escape maneuver was initiated before or at the point of peak outflow velocity, but there was no advantage in lateral maneuvering if the aircraft was already located after the microburst core.

- Lateral maneuvering reduced the advanced warning required to escape from microburst hazards. For example, in the simulation of a severe microburst encounter, a 10° lateral escape maneuver could be executed up to 38 sec later than the longitudinal-only maneuver with similar recovery performance.

- Although the difference was not significant, direct axisymmetric penetrations of the microburst core have found to be less hazardous than those which were slightly offset from the core.

- For trajectories offset from the microburst axis of symmetry, simulation of incorrect lateral maneuver (ie., turning towards the microburst core), resulted in increased performance loss.

- Turning away from the microburst core showed the greatest advantages in all cases.

In conclusion, if information about the location of the microburst core is available and the aircraft has not penetrated the microburst center yet, a lateral maneuver should be applied in order to turn away from the microburst core. If the aircraft is too close to the microburst core, high bank angles should not be used because insufficient reduction in the shear magnitude will be obtained to compensate for the performance loss associated with the turn. For the aircraft studied in this work, the optimal bank angle for average cases studied was 10°. If there are doubts concerning the location of the microburst center, an incorrect turn toward the core will be more hazardous than straight flight if the aircraft incorrect turn takes it through the microburst core. If the aircraft is far from the peak outflow velocity region, then risks of the incorrect turn are attenuated and the choice of lateral maneuvering or flying straight should depend on the judgement of the pilot based on the data he has available and visual clues. In this case, if the lateral maneuver is chosen, only low angles of bank should be applied. Finally, the absolute success of lateral maneuvering for microburst avoidance is limited to the availability of precise information about the existence and location of the microburst core.

Acknowledgments

Denise Avila de Melo was supported by CAPES (Federal Educational Agency of The Brazilian Government) under fellowship 27/87-2 and by Embraer S. A. (Brazilian Aeronautical Enterprise) during the development of this work as a Master of Science thesis research at the Massachusetts Institute of Technology. R. John Hansman was supported by the Federal Administration and the National Aeronautics and Space Administration under grants NGL-22-009-640 and NAG-1-690 and the MIT Lincoln Laboratory under contract BARR-10-119. The authors would like to thank Richard M. Hueschen from NASA Langley Research Center for sending the aircraft data and for his interest in helping with the research work.

References

1. U. S. Department of Transportation *Federal Aviation Administration Windshear Training Aid* Associate Administration for Development and Logistics, Washington, D. C., February 1987

2. Avila de Melo, Denise *Analysis of Aircraft Performance During Lateral Maneuvering for Microburst Avoidance* Master of Science Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, September 1989

3. Psiaki, M. L. and Stengel, R. F. *Analysis of Aircraft Control Strategies for Microburst Encounter* Journal of Guidance, Control and Dynamics, Vol. 8, No. 5, September-October 1985, page 553

4. Etkin, Bernard *Dynamics of Atmospheric Flight* John Wiley & Sons, Inc. 1972

5. Pines, S.; Schmidt, S. F. and Mann, F. *Automated Landing, Rollout and Turnoff Using MLS and Magnetic Cable Sensors* NASA CR-2907 National Aeronautics and Space Administration, Washington, D. C., October 1977

6. Oseguera, Rosa M. and Bowles, Roland L. *A Simple Analytic 3-Dimensional Downburst Model Based on Boundary Layer Stagnation Flow* NASA TM-100632, National Aeronautics and Space Administration Langley Research Center, Hampton, VA, July 1988

7. Miele, A., Wang, T., Melvin, W. W. *Guidance Strategies for Near-Optimum Take-off Performance in a Windshear* Journal of Optimization Theory and Applications, Vol. 50, No. 1, July 1986

8. McRuer, Duane; Ashkenas, Irving and Graham, Dunstan *Aircraft Dynamics and Automatic Control* Princeton University Press 1973

9. Åstrom, Karl Johan *A Simnon Tutorial* Department of Automatic Control, Lund Institute of Technology, Revised edition, July 1985

10. Dole, Charles E. *Flight Theory and Aerodynamics - A Practical guide for operational safety* John Wiley & Sons, inc. 1981

11. Frost, Walter and Bowles, Roland L. *Wind Shear Terms in The Equations of Aircraft Motion* Journal of Aircraft, Vol. 21, No. 11, November 1984

12. Kermode, A. C. *Mechanics of Flight* Longman Group UK Ltd 1987

13. National Research Council (U. S.) Committee on *Low-Altitude Wind Shear And Its Hazard to Aviation* National Academy Press, Washington, D. C. 1983

14. Rolfe, J. M. and Staples, K. J. *Flight Simulation* Cambridge University Press 1986

15. Wolfson, Marilyn M. *Characteristics of Microbursts in The Continental United States* The Lincoln Laboratory Journal, Vol. 1, No. 1 1988

Hazard Evaluation and Operational Cockpit Display of Ground-Measured Windshear Data

Craig Wanke* and R. John Hansman Jr.**
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, Massachusetts

Abstract

Information transfer issues associated with the dissemination of windshear alerts from the ground are studied. Two of these issues are specifically addressed: the effectiveness of different cockpit presentations of ground-measured information, and assessment of the windshear hazard from ground-based measurements. Information transfer and presentation issues have been explored through pilot surveys and a part-task Boeing 767 'glass cockpit' simulation. The survey produced an information base for study of crew-centered windshear alert design, while the part-task simulations provided useful data about modes of cockpit information presentation for both windshear alert and ATC clearance delivery. Graphical map displays have been observed to be exceptionally efficient for presentation of position-critical alerts, while some problems with text displays have been identified. Problems associated with hazard assessment of ground-measured windshear information are also identified.

1. Introduction

Low altitude windshear is the leading weather-related cause of fatal aviation accidents in the U.S. Since 1964, there have been 26 accidents attributed to windshear resulting in over 500 fatalities.[1,2] Low-altitude windshear can take several forms, including macroscopic forms such as cold-warm gustfronts down to the small, intense downdrafts known as microbursts. Microbursts are particularly dangerous and difficult to detect due to their small size, short duration, and occurrence under both heavy precipitation and virtually dry conditions. For these reasons, the real-time detection of windshear hazards is a very active field of research. As detection systems become operational, a host of operational issues arise. With the advent of digital ground-to-air datalinks and electronic flight instrumentation, there are many options for implementation of windshear alerts in the terminal area environment. Study is required to determine the best content, format, timing, and cockpit presentation of windshear alerts in the modern ATC environment in order to best inform the flight crew without significantly increasing crew workload.

2. Ground-Based Windshear Detection and Warning

2.1. Ground-Based Detection Technology

Ground-based windshear detection will play a large role in near-term windshear alerting and avoidance systems. The currently operational Low-Level Windshear Alert Systems (LLWAS), networks of anemometers around an airport and its approaches, are being expanded at some airports and are more capable of detecting windshears which impact the ground within the network boundaries. Much more importantly, ground-based doppler weather radar systems such as Terminal Doppler Weather Radar (TDWR) and doppler processing of ASR-9 radar data are becoming available and are capable of locating and measuring windshear events throughout the terminal area. These systems will provide the core data for near-term windshear alerting systems. Airborne predictive (look-ahead) sensors currently in the R&D phase - such as airborne doppler radars, lidars, and infrared detectors - will supplement ground-derived data as they become

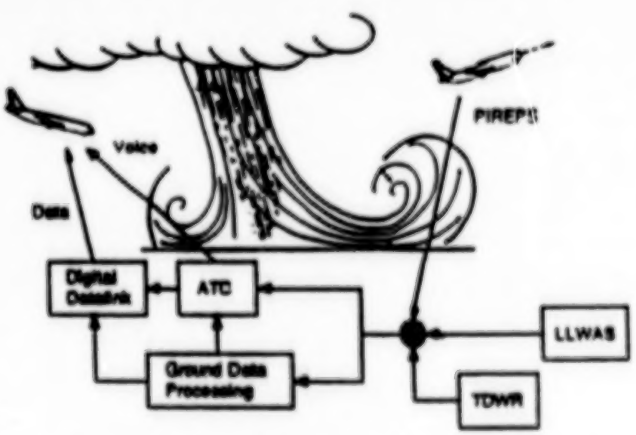


Fig. 1: Integrated Ground-Based Windshear Alerting System

---

* Research Assistant, Student Member AIAA
** Associate Professor, Associate Fellow AIAA

Copyright © 1989 by M.I.T. Published by the American Institute of Aeronautics and Astronautics, Inc. with permission.

available and economically feasible. Airborne *in-situ* or reactive windshear sensing, provided through comparison of airspeed measurements with inertial accelerometer measurements, is a currently available technology which can provide flight crews with a warning of windshear penetration.

2.2. Integrated Ground-Based Systems

Assuming the near-term (early 1990's) deployment of both ground-based doppler weather radars and the Mode-S ground-to-air digital datalink, possible paths of information flow are illustrated in Figure 1. In this environment, data from LLWAS and TDWR sensors can be combined with pilot reports (PIREPs) to form the current windshear database. These PIREPs may be verbal, or reported automatically by an airborne inertial sensor over the digital datalink. This data can then be processed to varying degrees, and transmitted to the aircraft via voice communications or digital datalink. Several issues are raised by this implementation. One of these is the degree of data processing done; this can range from transmission of essentially raw data (as in the original LLWAS implementation, for example) or complete processing of the data into an executive decision to close the runway. One consideration is purely operational; what should be the distribution of decision-making responsibility between the pilot and the ATC controller? Another consideration is technical; given the available weather information, what is the (quantitative) hazard posed by the current weather situation to a particular aircraft or aircraft type?

Another of these issues is the "crew interface", the procedure and method used to inform the crew of a hazard. An essential difficulty in presenting windshear information is the need for alerts during descent and final approach, which are high workload phases of flight. For this reason, design of the crew interface is critical; a poor interface will result in loss of information or increased crew workload. The advent of electronic cockpits and digital ground-to-air datalink opens up a variety of options for implementation of the crew interface. Some issues to be examined include: information content, message format, and mode of presentation.

The information content of all uplinked alerts is limited by either message length constraints on digital information or the limited time available on voice channels. This forces a tradeoff between information content and message length, requiring determination of which information items are most essential to crew comprehension. If a voice channel is to be used, the format of the message must be designed such that the information is perceived with the intended urgency and does not get lost in the large amount of voice traffic in the terminal area. If the information is sent via datalink, the data must be presented in an easily understood manner, so as to minimize the time required for comprehension of the alert while still conveying all the necessary information. Uplinked messages can take advantage of the electronic instrumentation in modern transport-category aircraft, such as moving map displays, to meet these requirements.

2.3. Research Focus

The specific focus of this research has been the evaluation, transmission, and presentation of ground-based doppler weather radar derived information through a limited bandwidth digital datalink (Mode-S). The first issue studied is the content and cockpit presentation of uplinked windshear alerts. User input was solicited through pilot opinion surveys, and then used to design a part-task simulation experiment. The primary results deal with the use of electronic instrumentation for presentation of uplinked information; specifically the relative merits and disadvantages of voice, alphanumeric (textual), and graphical modes of presentation. In this context, *voice* or *verbal* mode refers to standard ATC radio communications, *alphanumeric* or *textual* mode refers to presentation (on some electronic or paper device) of the literal text of a message, while *graphical* mode refers to a combined pictorial/text presentation of the alert information on some electronic map or map-like display. Alphanumeric and graphical presentations presuppose the existence of a ground-to-air digital datalink.

The second issue examined is the evaluation of ground-measured windshear data to determine a hazard index. This hazard index should both accurately quantify the windshear hazard present and be meaningful to the flight crew. Overwarning must be minimized, since a large number of false or nuisance alerts can disrupt airport operations and damage pilot confidence in the alerting system. Preliminary analysis has identified some of the issues and problems involved, and further work is in progress.

3. Crew Interface Issues

3.1. Pilot Opinion Surveys

In order to obtain user input on both current windshear alert systems and requirements for future systems, a pilot opinion survey has been conducted. With the cooperation of the Airline Pilots Association and United Airlines, responses have been collected from 51 line pilots of transport category aircraft with autoflight systems (Boeing 757, 767, 747-400). Significantly, 51% of the pilots have had what they considered to be a hazardous windshear encounter; most of these occurred at Denver-Stapelton airport, a UAL hub and an area noted for heavy microburst activity during the summer months. It should also be noted, however, that pilots who have had a hazardous

windshear encounter may have been more likely to respond to the survey. Some general results:

- Most of the pilots (90%) agreed that "Microbursts pose a major safety hazard to transport category aircraft."
- Only 15% of the respondents agreed that "Currently available windshear alert data is sufficient for safe operation in the terminal area," while 44% disagreed.
- *All but one* (98%) of the pilots felt that "a system to provide aircrews with better and more timely windshear alerts is necessary."

These responses clearly indicate that pilots are dissatisfied with current windshear alert data and would be receptive to improvements. The pilots were also asked to rate the usefulness of currently available windshear data (Figure 2). Significantly, PIREPs and visual clues are both considered more useful for windshear avoidance than LLWAS. However, neither PIREPs or visual information are always available; this emphasizes the need for a remote detection and advance warning system.

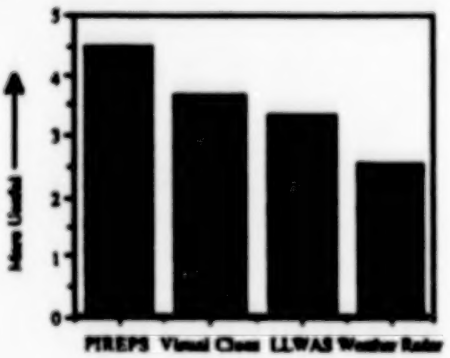


Fig. 2: Pilot ranking of windshear information sources

Some of the survey questions were designed to obtain information about user needs and preferences for use in the design of windshear alert presentations. These questions assume the existence of a ground-to-air datalink and an Electronic Flight Instrumentation System (EFIS). Design issues for this scenario include decisions about both message content and mode of presentation. Because of the high workload in terminal area operations, it is important to consider the manner in which information is presented to the flight crew. In the integrated ground-based system considered in Section 2, there are three modes of information presentation available in a modern cockpit: *verbal*, *alphanumeric*, and *graphical*. These modes have been defined in Section 2.3. Issues to be considered include crew workload, preferences, and the capabilities of the aircraft instrumentation.

Responses from the pilot survey indicate that pilots are receptive to graphic displays. (Figure 3). The specific suggestion of integrating windshear information with an EFIS moving map display was strongly supported. Also of interest was the high preference for ATC voice alerts, which is likely a result of a practiced ability to interpret radio communications. Display of windshear alerts on some alternate graphical display (other than the EFIS moving map) was also ranked above alphanumeric displays and ATIS. Comments received indicated that the low ranking of ATIS was due to the long time between updates.

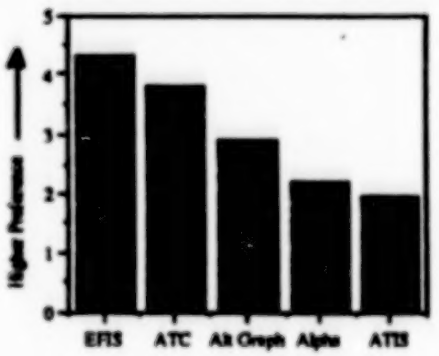


Fig. 3: Pilot rankings of possible presentation modes for ground-generated windshear alerts

Due to voice channel time and digital bandwidth constraints, the amount of information space available for a given alert is limited. For this reason, message content is critical. A question dealing with the message content of microburst alerts was included in the pilot surveys. The responses indicate that location and intensity of microbursts are clearly the most important information items. Size, microburst movement, and intensity trends are of secondary importance, and shape data is generally felt to be least important. Ranking of this information allows the design of alerts which fit within the message length constraints and still retain enough relevant information to be useful. In this case, the data indicates that a message composed of location, intensity, and perhaps size would be sufficient.

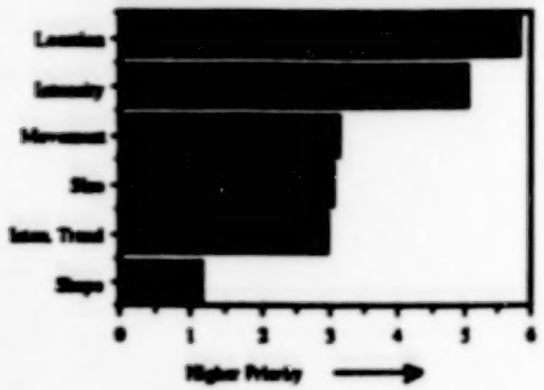


Fig. 4: Pilot ranking of microburst information by importance

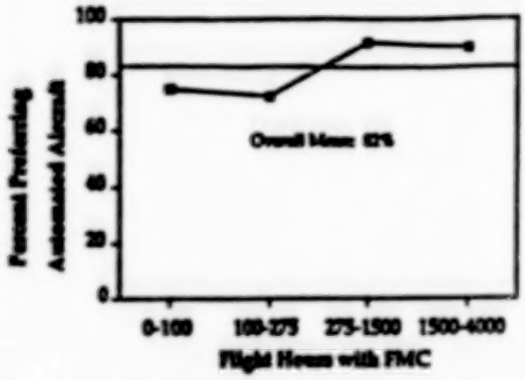
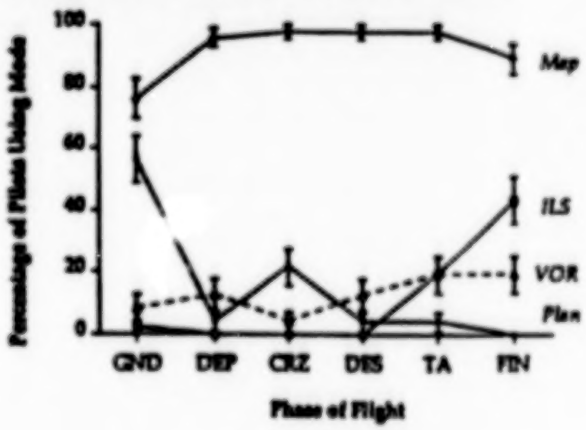


Fig. 5: Crew Preference for Automated Aircraft [3]

The survey also addressed timing of microburst alerts and threshold shear levels. There was no consensus as to in what phase of flight (during the approach) alerts should be given; the most common response was "as soon as detected." The respect with which pilots treat the windshear threat was emphasized in a question about threshold shear levels. The average response was that about 10 knots of headwind-to-tailwind shear component (i.e. airspeed loss) should constitute a windshear advisory and only 15 knots a windshear warning. Also, it was almost unanimously expressed that decisions about the threat posed by a particular windshear situation should be made entirely by the pilot, and the controller's role should be to maintain safe separation during avoidance maneuvers. These responses all indicate that the pilots would like to have all the information available as soon as measured, and the sole responsibility for evaluation of a particular hazard situation. Due to crew and ATC saturation problems, it is impractical to plan on distribution of all available windshear information to all aircraft in a congested terminal area. Therefore, a uniform hazard assessment criterion needs to be set for ground-measured information and a corresponding threshold defined for which the threat becomes signif[illegible].

Because all of the flight crews included in this survey were qualified on autoflight aircraft, specific questions on use of cockpit automation systems were included. This section was specifically concerned with use of a Flight Management Computer (FMC) in concert with an EFIS, and was intended to evaluate crew acceptance and usage patterns of these automated systems. In general, regardless of flight hours with the FMC, pilots expressed a decided preference for automated aircraft over non-automated ones (Figure 5). Also, the consensus was that the FMC significantly reduces workload in most phases of flight, with the exception of the pre-flight programming required.[3]

The crew use pattern of the Electronic Horizontal Situation Indicator (EHSI) was also addressed. The EHSI is a map-like display of the aircraft's currently programmed flight path in addition to weather and other navigational information. The data, illustrated in Figure 6, demonstrates that the crews use the moving map mode of this display almost exclusively in almost all phases of flight. The exceptions are use of Plan (north-up) mode during ground operations for flight path programming and some use of the ILS mode during final approach. Since the simulation scenarios are set during descent and approach, Map and ILS modes were included in the EHSI simulation.



Phases of Flight:

GND = Ground Operations	DES = Descent
DEP = Departure	TA = Terminal Area
CRZ = Cruise	FIN = Final Approach

Fig. 6: Use of EHSI Modes by Phase of Flight [3]

An effort was also made to determine the information density of the EHSI in Map mode. Figure 7 indicates that the "information load" on the EHSI, calculated by averaging pilot need ratings for the discrete items shown on the map display (see Figure 9).

peaks during descent and terminal area operations. This indicates a potential clutter or information overload problem for alphanumeric or graphical alerts which use the map display, re-emphasizing the need for a well-designed alert system.

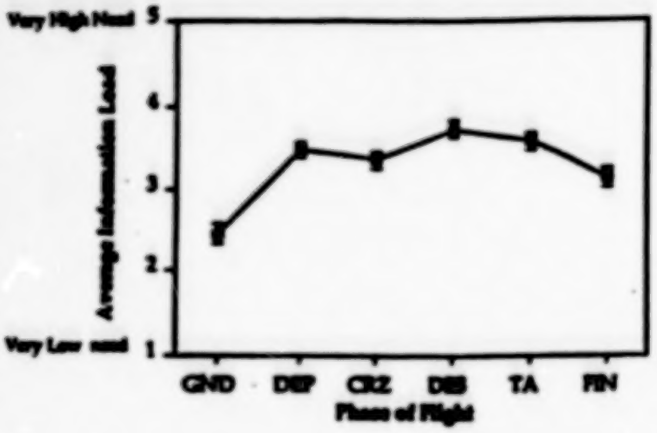


Fig. 7: EHSI Information Load by Phase of Flight [3]

3.2. Flight Simulator Study

3.2.1. Simulator Design: Hardware

The purpose of the part-task simulator was to allow comparison of presentation modes - *verbal*, *alphanumeric*, and *graphical* as defined in Section 2.3 - on a modern 'glass cockpit' aircraft. The Boeing 757/767 class of aircraft with its Electronic Flight Instrumentation System (EFIS) has been simulated. This simulator (Figure 8) contains the following elements:

Electronic Flight Instrumentation System (EFIS) and other instruments:

- EADI - Electronic Attitude Director Indicator: artificial horizon, autopilot annunciations, groundspeed, radio altitude
- EHSI - Electronic Horizontal Situation Indicator: moving map display with either track-up MAP mode with programmed path and navigational information, or heading-up ILS mode with glideslope and localizer needles. Both modes can be operated in 6 different ranges and permit overlay of airborne weather radar (WXR) reflectivity information. The EHSI was also used for graphical microburst windshear alerts. (Figure 9).
- Airspeed Indicator, Altimeter, and Vertical Speed Indicator: electronic "moving tape" displays of these instruments.
- Marker Beacon indicators, Flap indicator dial, Gear lights
- Alphanumeric display window, used for display of alphanumeric windshear alerts
- Sidetask display: A simple meter and buttons were displayed below the EHSI to provide a mouse-driven following sidetask for workload monitoring

Flight Management Computer (FMC):

- CDU - Control Display Unit: an alphanumeric display and keyboard for pilot input and control of the FMC. The simulator CDU has "pages" (display screens) for route legs (horizontal and vertical path programming), direct-to (for flight directly to a fix or waypoint), and for setting the intended landing airport and runway.

Control Panels:

- Autopilot Glareshield Panel: a simulation of the controls for the 757/767 autothrottle and autoflight systems, including LNAV/VNAV flight (following FMC-programmed lateral and vertical paths) and the various capture and hold modes for airspeed, heading, vertical speed, and altitude guidance.
- EHSI Panel: allows setting of the map display range from 10 to 320 nm, switching between MAP and ILS modes, and suppression of WXR, navaid, intersection, or airport information if desired.
- Flaps and Gear Panel: includes a switch for landing gear and a rotary dial for flap setting.

Communications:

- Headsets were provided for the pilot and simulation controller for verbal ATC communications; this system also provided a switch which the controller used to trigger some simulation events.

The hardware required to implement this system consists of an IRIS graphics workstation for the electronic displays and calculation of flight dynamics and auto-navigation, an IBM-XT computer for the flight management system, and a set of three control panel boxes with data acquisition hardware.

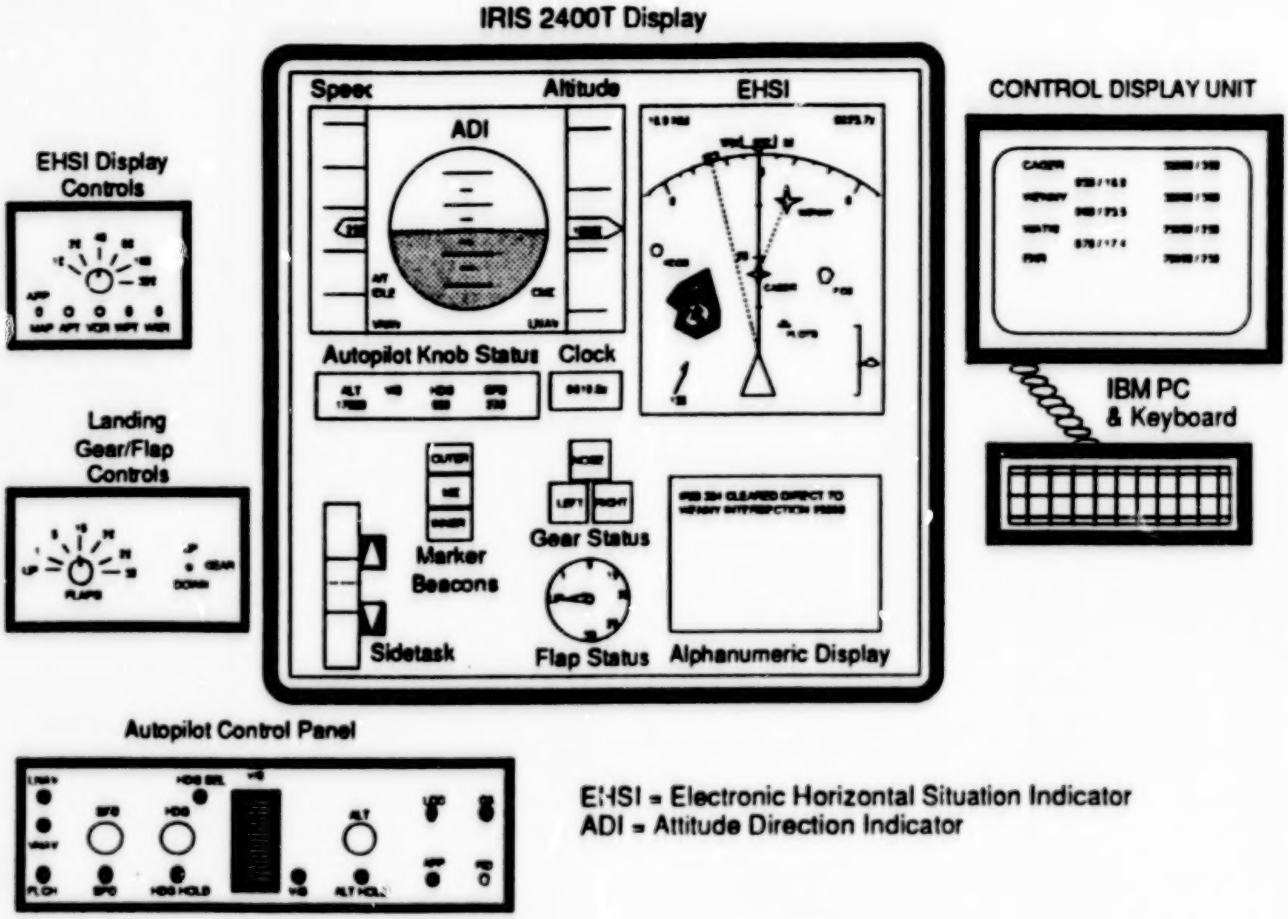


Fig. 8: Boeing 757/767 part-task simulator

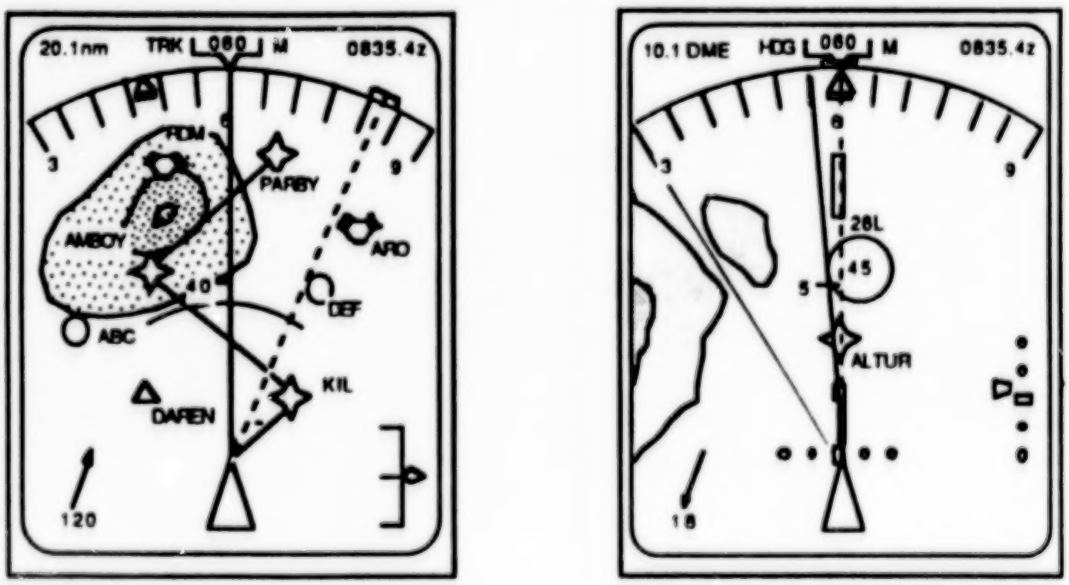


Fig. 9: Boeing 757/767 EHSI display modes: MAP (left) and ILS (right)

This experiment was concerned with cognitive decision-making issues rather than the details of pilot performance. This allowed the exclusion of controls and instruments not related to the particular cognitive task at hand. The lack of a copilot and imposition of a sidetask compensated for the workload loss. The subjects generally agreed that the simulation was accurate for the tasks they were asked to perform. Also, no windshear dynamics were included, in that the data of interest was the go/no-go decision and whether

or not penetration occurred. The major advantages of the part-task simulator are the ease of setup and operation and the flexibility of the electronic displays. Alphanumeric and graphical message formats are easy to implement and change.

3.2.2. Scenario Design

The initial set of experiments was designed to answer the question: what are the advantages and disadvantages of graphical, alphanumeric, and verbal presentations in the context of both windshear alerts and ATC amendments? To accomplish this goal, a set of nine descent and approach scenarios into Denver-Stapelton airport was devised. The selection of the Denver terminal area is advantageous for two reasons: 1) the high incidence of dry microburst activity observed there and 2) the large number of descent profile and landing runway combinations possible. The inclusion of both ATC amendments and microburst alerts into the same scenario was useful in preventing the subject from anticipating or overreacting to repeated windshear alerts.

Each of the scenarios was divided into two phases. The aircraft was initially positioned at the outer limit of the terminal area, and given an initial flight plan (pre-programmed into the FMC). During the descent phase, the pilot was given three clearance amendments which required reprogramming of the FMC for compliance.

The second phase of the scenario began when the aircraft was vectored onto the final approach course. At this point, windshear alerts could occur. Microbursts were positioned either as a threat on the approach path or as a non-threat on the approach or departure end of another runway. In addition, microbursts could be positioned on the missed approach path. The alert was given either close in at the outer marker (6 to 9 nm from touchdown) or further out at 20 nm with a second message at 10 nm from the runway threshold.

The nine generic scenarios were divided into three blocks of three by presentation mode. In each block, all amendments and windshear alerts were given in the assigned mode: verbal, alphanumeric, or graphical. Verbal clearance amendments were given according to current ATC operating procedure. Alphanumeric clearance amendments were activated remotely by the controller, generating an audible alert, and the text of the message appeared on the CDU(Control Display Unit: the input screen for the FMC). In the graphical mode, clearance amendments when activated appeared directly on the EHSI as an alternate route (dashed white line), and could be accepted or rejected with a single FMC keystroke.

Microburst alerts always contained warnings for all possible approach runways, not only the one being used by the approaching aircraft. This was to ensure that all modes had the same information content, and to measure the pilot's facility to determine threat from non-threat situations in all three modes. Verbal microburst alerts were given as standalone messages by the controller. Textual microburst alerts appeared in an alphanumeric window just beneath the EHSI display. A typical verbal or textual alert: "IRIS 354, Microburst Alert. Expect four-zero knot loss, 2 mile final approach runway one-seven-left." Graphical microburst alerts appeared in the appropriate location on the EHSI map (in either MAP or ILS mode) as flashing white circles with the intensity (headwind-to-tailwind divergence value in knots) drawn over them. An example is shown on the ILS mode display in Figure 9. Verbal cues were given (i.e. "IRIS 354, Microburst alert.") in all modes, so that the time of notification was kept constant; this would not be the case in an actual cockpit, where an automated audible alert would most likely be used. Over the subjects tested, all scenario blocks were tested in all the modes, and the order in which the subject encountered the modes was rotated. This process was used to eliminate learning and scenario-dependent effects.

3.2.3. Experimental Procedure

At the start of the session, the pilot was given all of the appropriate charts for the Denver-Stapelton area, the initial clearances for the nine scenarios, and the required checklists. He was then asked to complete the first stage of a NASA-designed workload evaluation[4], which asked him to prioritize the different types of workload for the specific task of flying a 757/767 aircraft. At this point, the features of the simulator were demonstrated, and a sample scenario flown in which all of the three modes were encountered for both phases of flight. After all the subject's questions were answered, the testing began. Each of the nine scenarios lasted from 20 to 35 minutes. During the flights, one of the experimenters served as the ATC controller and one remained in the cockpit with the pilot to answer questions about physical operation of the simulator. After each scenario the pilot completed an evaluation sheet in which he described what level of workload he felt he was under, as well as his level of performance. After the scenarios were all complete, there was a debriefing session in which the pilot's impressions of both the simulator and the presentation modes tested were solicited. The data taken included:

- Computer recording of flight data, control inputs, sidetask performance, and FMC programming

- Audio tapes of ATC communications and 'cockpit' voices
- Videotape of EFIS displays
- NASA subjective workload evaluation sheets
- Post-session debriefing

From this data was determined the percentage of "correct decisions", time taken to make decisions, level of workload, and a number of qualitative observations. An "incorrect decision" was scored for either microburst penetration or a missed approach in response to a non-threatening microburst alert.

3.2.4. Results

The major quantitative results are illustrated in Figures 10 through 12. It was found from all three figures of merit - performance, workload, and preference - that the graphical presentation mode was superior. Also, the textual (alphanumeric) presentation mode proved roughly equal or slightly inferior to standard verbal communications. This data set was based on runs with eight 757/767-qualified active line pilots, all with considerable FMC experience.

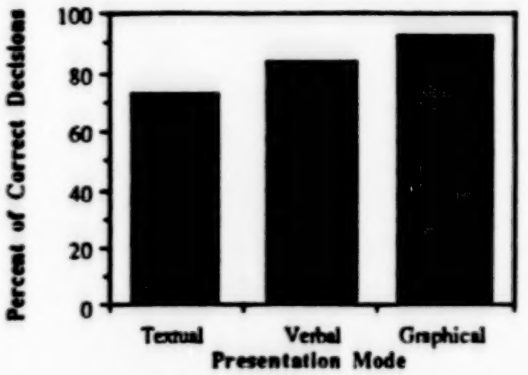


Fig. 10: Part-task simulation results - response performance to windshear alerts by mode

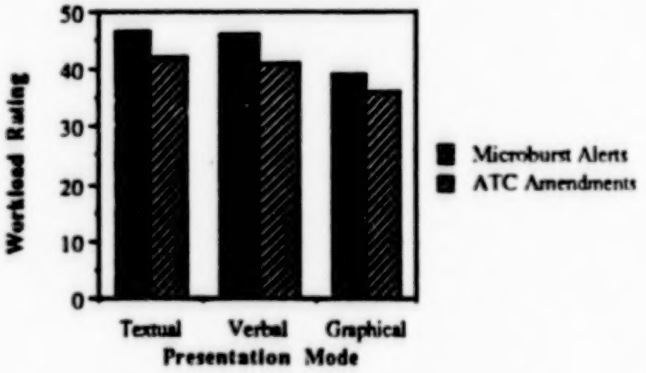


Fig. 11: Part-task simulation results - subject workload by mode

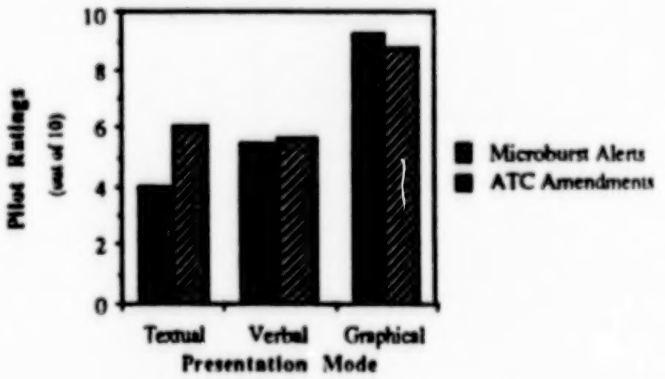


Fig. 12: Part-task simulation results - pilot preference by mode

These numerical data were interpreted based on qualitative observations and pilot comments. In general, the graphical mode was most effective due to short comprehension time. Placing the data on the EHSI map display allowed the pilot to quickly visualize the situation, and spend less time orienting himself to the change in conditions resulting from the amendment or alert. Information on the EHSI seems to be more consistent with the cognitive map formed by the pilot. The rapid comprehension also meant a minimum of added "head-down" time, which appealed to the pilots. In the case of microburst alerts, the positional information contained in the graphical mode actually led several pilots to request and program non-standard missed approach procedures in advance in order to avoid the windshear areas completely. When the pilots were given the same information in the other modes, this was generally not observed.

Evaluation of the results for textual and verbal modes was more difficult. One advantage of the textual mode for amendment delivery is the elimination of copying errors, especially with 5-letter intersection names. For windshear alerts, which are not copied down, more erroneous decisions are made with the textual than with the verbal mode. This implies that the information, given during the time-critical final approach phase, is more prone to misinterpretation

consistent with the low preference rating expressed by the pilots for the textual microburst alerts. In most cases the pilots indicated that they did not like to look away from the instruments to read the textual alerts, most likely leading to poor comprehension. An additional problem with textual alerts is due to vision; textual alerts require refocussing, which was observed to be more difficult for older pilots.

Verbal alerts also scored lower workload ratings for both experiments, and the performance advantage of verbal over textual for the windshear alerts was fairly large. It is apparent that the pilots had a great deal of experience in comprehending and retaining verbally transmitted information. For this reason, even though the verbal and textual modes contained exactly the same information, and the textual information stayed visible for several minutes, the verbal alerts were more effective. Some observations taken from pilot comments:

- Textual alerts in time-critical situations - such as final approach - require too much head-down time.
- Digitally transmitted information in either mode, textual or graphical, leads to a loss of prosodic (voice inflection) information. Since controllers sometimes use voice inflection to distinguish urgent alerts from normal communications, this is in some sense a loss of information.
- Digitally transmitted information, if directed to specific aircraft, prevents pilots from hearing instructions given to other aircraft in the terminal area. Some pilots stated that hearing the communications to other aircraft in the vicinity gave them a better understanding of the overall situation and enabled them to be better prepared when an alert arrived. Other pilots indicated that they could do without the information.
- Increasing the level of automation in the cockpit may lead to a lax attitude on the part of the flight crew, possibly leading to a loss of situational awareness and slow reactions in the case of an emergency.

A final observation that has implications for the design of cockpit data presentation is that many pilots took the most time-critical portion of a message, resolved it, and then went on to complete the task when the time pressure was off. This argues for splitting long or complex messages into two on more parts for transmission and/or presentation to improve the pilot's comprehension speed.

3.3. Crew Interface Research: Summary

The consistency between the survey and simulation results allows the following conclusions to be drawn. Pilots appear to be generally receptive to the idea of automated ground-to-air information transfer for both windshear alerts and clearance delivery. However, the presentation of the information is critical. From the pilot opinion survey results, the post-simulation interviews (pilot preferences), and the simulation results (workload, performance) it is clear that graphical presentation of both windshear alerts and clearance amendments is significantly more effective than verbal communications. This may possibly be generalized to any situation which requires the pilot to recognize and interpret spatial information quickly; the graphical presentation mode seems to allow much quicker comprehension of such information. To obtain this benefit, the detailed format of such graphical information must be carefully designed to present only the necessary information in clear fashion without clutter or data overload.

In the case of windshear alerts, the pilots identified this minimum presentation to be a simple symbol showing location, approximate size, and intensity. The proposed Mode-S datalink, for example, allows 48 bits of useful information every 4 to 12 seconds in surveillance mode. This minimum alert presentation can likely be expressed in 24 bits or less, allowing two messages per scan. Therefore, the Mode-S link can possibly be used to display and track several microbursts, while keeping up with the 1 minute update rate achieved by TDWR in the current configuration.

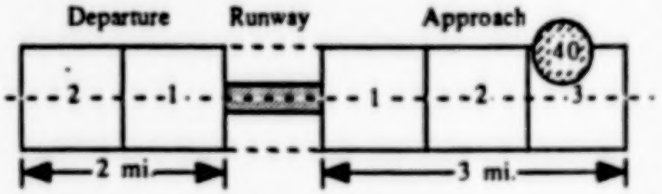
Information received over a digital datalink may also be presented as alphanumeric (textual) messages; again, the survey and the simulations were consistent in the results. The textual mode of presentation was rated poorly by the pilots, and did not effect either a reduction of workload or an improvement in performance. Pilots disliked in particular the additional head-down time required to read textual information. The speed of comprehension did not seem to improve with textual warnings; the familiarity of operational pilots with verbal communications seemed to outweigh the advantage of having the text of the message displayed indefinitely. However, spelling and pronunciation errors were eliminated with the text display. Finally, textual messages appear to offer no advantage over verbal messages for comprehension of spatial information.

4. Hazard Assessment

The successful implementation of ground-measured windshear alerts requires an effective way of quantifying

the windshear hazard. This hazard criterion must provide an accurate estimate of the danger to approaching and departing aircraft which can be easily interpreted by the flight crew. In order to maintain pilot confidence in a ground-based windshear avoidance system, an alert must correlate with what the aircraft is experiencing or will experience. Otherwise, even an accurate measurement will be perceived as a false alarm.

The method used in the recent TDWR operational evaluations at Denver-Stapelton and Kansas City International airports is based on horizontal wind measurements. When a change in low-altitude horizontal winds above a threshold is measured (radial from the radar), the area is marked and quantified by the maximum radial shear measured within it. If this area is within the segments identified in Figure 13, an alert is issued.[6,7]



The alert corresponding to the 40 knot microburst pictured above might be:"United 226, Denver tower, threshold wind one six zero at six, expect a forty knot loss on three mile final."

Fig. 13: Windshear Reporting Zones for Approach and Departure: 1988 TDWR Operational Evaluation

One problem with this system is that a microburst which occurs in one of the boxes may in one case never encroach on the flight path, and in another be right on the center of it. In either case, the same alert is issued. This means that approaching or departing aircraft may fly through the center of a microburst or almost entirely miss it, experiencing the entire event or nothing at all. This could be perceived as a false alarm by the pilot, although the windshear is present and may even be fairly close to his position. A recent report about the 1988 TDWR Operational Evaluation indicates that this could be a major issue.[8] Data was collected from 111 pilots who landed or took off during alert periods.* Of this group, 34% indicated that 'nothing was encountered', while another 31% reported something like 'nothing much was encountered'. A "nuisance alarm" rate this high can unnecessarily disrupt airport operations as well as damaging pilot confidence in the windshear alert system.

Another spatial problem relates to altitude of penetration. Microburst windfields, due to their small size and dynamic nature, can vary strongly over the lowest 1000 feet above ground level. With the current alert methodology, the reported airspeed loss will be the same for a microburst occurring one, two, or three miles before touchdown. However, the winds experienced along the approach will be much different. Using an analytic microburst model[9], the winds experienced by an aircraft on final approach for microburst encounters at 1, 2, and 3 miles from the threshold are plotted in Figure 14.** Changing the point of encounter, which corresponds to a change in penetration altitude, produces large changes in the actual windshear experienced.

This altitude variability is further aggravated by measurement limitations. A radar scan beam used to detect microbursts has a finite beamwidth on the order of 1°. For a radar situated several miles from the airport, this means that this scan effectively measures an average radial wind velocity over the lowest 500 to 1000 feet. As observed above, microburst windfields can vary strongly over the lowest 1000 feet. This indicates that some possibly important information is lost. Some correction factor for altitude based on a 'typical' microburst windfield might improve the fidelity of the warnings.

A third problem with these alerts is the intensity measurement used. The windshear quantity used in the alerts is derived from the maximum change in radial velocity over the area of shear. This number is reported in the alert as an X "...knot airspeed loss." In reality, for a roughly axisymmetric microburst, this number represents an X/2 airspeed gain followed by an X knot loss. This is not a reporting error, but crews should be aware that the quantity being measured is the total horizontal divergence over the shear, and that the maximum airspeed loss relative to the reference airspeed before penetration is actually one-half of the reported value.

---

* Since the microburst alerting software changed over the course of the 1988 Operational Evaluation, the pilot responses included in this calculation were those pilots who not only were alerted but who would have been also alerted by the final version of the warning algorithm. The earlier version of the algorithm produced more alerts than the final one.

** The parameters used for this case were taken approximately from the sample case given in the Oseguera and Bowles paper.[9] The values are: Maximum outflow velocity = 40 knots, Altitude of maximum outflow = 120 feet, Radius of downburst shaft = 2500 feet. It is important to note that altering these parameters can significantly change the winds experienced.

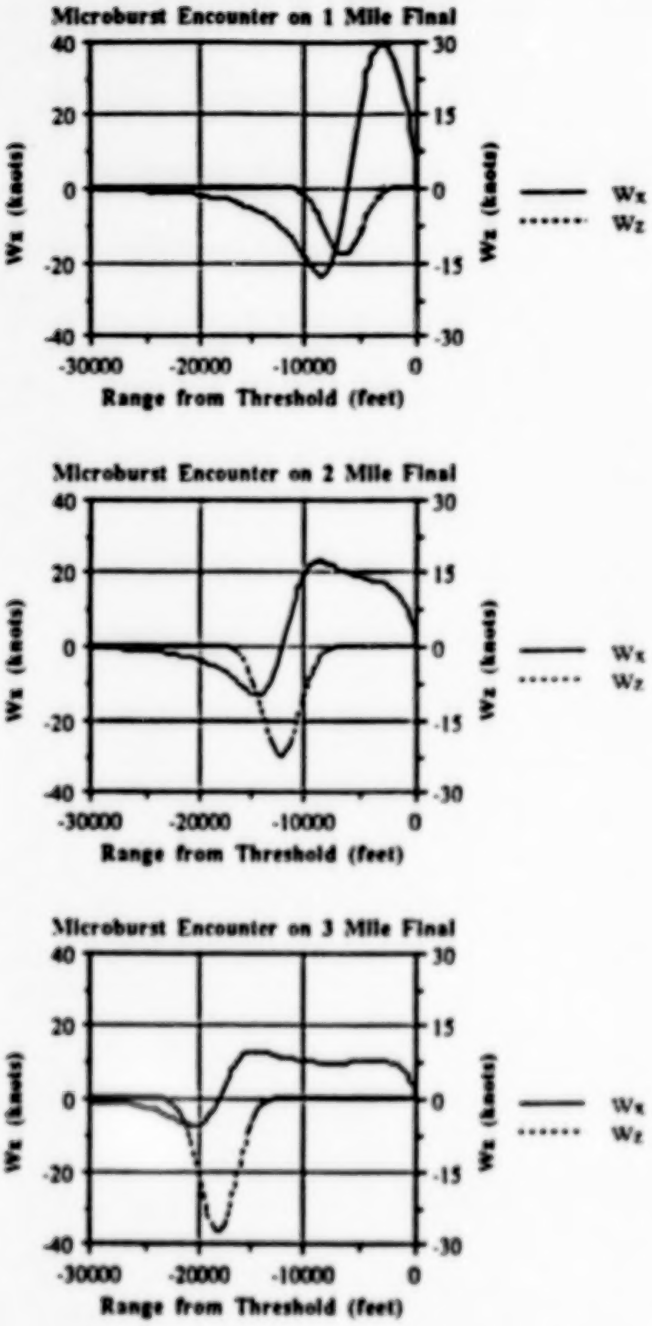


Fig. 14: Winds Experienced on Final Approach for a Microburst Encounter at Three Locations

The actual value of airspeed change experienced by a particular aircraft is also dependent strongly on the guidance strategy being employed by the pilot or autopilot, and to some extent on the spatial extent of the microburst. If the windshear occurs over a short distance, even though the airspeed fluctuations may be large, the deviation from glideslope will be small. If the shear has a large component at the phugoid frequency of the aircraft, and no strong corrective action is taken (essentially stick-fixed conditions), a small magnitude shear can cause large deviation from the glideslope. In an autopilot coupled approach, the effects of these conditions will be again different. In short, the dynamic characteristics of the pilot/autopilot/aircraft system significantly affect the impact of a certain shear on a particular aircraft.

One further measurement difficulty which need be noted is due to microburst asymmetry. The estimation of winds along the flight path by a doppler radar is currently done assuming axisymmetric microbursts, since the radar scan is in general not oriented directly along the flight path. Data shows that microbursts are frequently highly asymmetric, and that techniques for detecting this asymmetry could improve estimation of flight path winds considerably.[10]

The difficulty in defining a suitable measure of windshear intensity is due to the nature and limitations of the ground-based measurements. Knowledge of the windfield in the earth-fixed reference frame is different from the measurements taken by an airborne windshear sensor. An airborne sensor is basically confined to taking point measurements along the flight path of the aircraft; however, the aircraft state is known and can be incorporated into the warning. The currently accepted way of measuring a hazard based on this information is the 'F-factor', which measures the instantaneous loss of rate-of-climb due to the windfield immediately ahead.[5] This measurement is indicative of the loss of escape capability of the aircraft, should a windshear avoidance maneuver be necessary. In this case, a warning is generated when F exceeds a threshold value for a predetermined length of time.

Ground-based measurements do not have such a complete knowledge of aircraft state, and it would be difficult to derive F from such a measurement. However, a ground-based doppler weather radar measurement can encompass the entire windshear area and be used to make an assessment of the overall situation, which is a great advantage. The hazard due to a particular microburst event is strongly dependent on the spatial effects (off-center microburst penetration, altitude of penetration, or microburst radius) detailed above. The advantage is that these effects can be measured and incorporated into a ground-generated windshear alert.

The problem is then one of incorporating some or all of these effects into the alert and/or the intensity measurement to minimize pilot-perceived false alarms or "nuisance alerts." Some of this could be accomplished by modifications of the alert format. One such change (proposed in Reference 8) is to add the words 'left', 'right', or 'center' to the alert to indicate the microburst position relative to the flight path. Another approach is to use correction factors for some of these effects. These can perhaps be developed from flight simulation studies using both analytical and dual-doppler derived microburst windfields. One difficulty with a flight simulation study is the definition of a baseline flight condition. For example, should the hazard assessment represent a worst case scenario (stick-fixed response, no escape attempted) or include the damping effects of a 'typical' pilot or autopilot glideslope tracking technique? This would also differ for approaching and departing aircraft. Also, a measure of flight path deterioration needs to be defined. Maximum excursion below glidepath (in degrees or feet) is one possibility.

In addition to a good hazard assessment criterion, it is important that crews be better informed about the workings of the system. In particular, the actual meaning of the alert should be made more clear. The possibility of a microburst being to the side of the flight path should be expressed, and the real meaning of the windshear divergence value should be explained. Crews should also be aware of the measurement limitations of the sensing system.

5. Conclusions

- A pilot opinion survey and a flight simulator experiment have been performed in order to examine issues related to dissemination of ground-measured windshear information to flight crews with and without a digital datalink.

- Survey results indicated that the currently available windshear avoidance information is not sufficient, and that a better system is highly desirable. A preference for graphically presented microburst alerts was expressed, and some specific questions about the makeup and timing of microburst alerts were answered. The survey results were then used in design of the flight simulator experiment.

- Simulation experimental results indicate that presentation of windshear alerts as graphical symbols on a moving-map display is significantly more effective than verbal alerts. Pilot performance improved, and pilot workload decreased. Both the survey results and comments from the simulation subjects indicate a strong pilot preference for graphical presentations. It is believed that the map representation is more consistent with the pilot's cognitive map and that graphical information is therefore more quickly and accurately assimilated.

- Presentation of windshear alerts as text on an electronic display proved inferior to standard verbal communications in terms of workload increase, pilot performance, and pilot preference. The survey respondents indicated that too much 'head-down' time is required to read text messages during final approach; this was corroborated by the simulation subjects. In time-critical situations (i.e. microburst alerts) it was apparent that textual messages were more subject to misinterpretation than verbal ones. In the non time-critical case of clearance amendments, textual messages did have the advantage over verbal messages of eliminating copying errors; however, no significant improvement in performance or reduction in workload was observed. In either case, the familiarity of pilots with verbal communications allowed them to comprehend the message quicker than in the textual mode.

- Some more general cockpit display design considerations were observed in this study. The use of an aircraft-directed digital datalink such as Mode-S allows more sophisticated information presentation, but deprives the flight crew of listening to transmissions to other aircraft as well as prosodic (voice-inflection) information from the controller. Also observed in the simulations was the fairly universal and consistent practice of separating information into time-critical and non-time-critical pieces. This could be used to design displays or message formats which emphasize certain parts of a message to eliminate the time spent prioritizing the information. Another issue discussed by the simulation subjects is a possible decrease in crew alertness due to increased cockpit automation.

- It is critical to the implementation of a ground-based windshear alerting system to quantify the windshear hazard both accurately and clearly. Overwarning can unnecessarily disrupt airport operations as well as damaging pilot confidence in the windshear alert system. The system used in the 1988 TDWR Operational Evaluation was shown to result in a significant number of "nuisance alerts," for which the pilots reported experiencing little or no windshear.[8] To address this problem, it is proposed that 1) a better

method of assessing the windshear hazard be developed, and that 2) flight crews be better educated about the meaning of ground-generated windshear alerts.

- The assessment of the windshear hazard from ground-based doppler radar measurements is different than from airborne measurements. The ground-measured data covers a larger area around the aircraft flight path, but lacks specific information about aircraft state. The problems involved in such a hazard assessment can be split into two categories: quantification of the windshear hazard from the knowledge of the wind velocities and microburst locations in the earth-fixed reference frame, and consideration for the limitations of ground-based doppler radar measurements. Problems in the first category include spatial effects (such as off-center microburst penetration and altitude of penetration), and aircraft-dependent variations. Problems in the second category include finite radar beamwidth, the inability to accurately measure the vertical velocity field, and microburst asymmetry. Through analysis and flight simulation, an improved hazard assessment criterion can be developed.

- Educating flight crews about the meaning of ground-generated alerts is equally as important as a good hazard criterion. The microburst alerts currently used, for example, indicate the maximum measured horizontal wind change and the approximate location. Pilots should be informed that the microburst as reported may lie to one side of the approach or departure flight path, and for this reason little or no windshear may be experienced. In addition, windshear magnitude can vary strongly with altitude of penetration. Also, the value measured is currently given as an "airspeed loss." This actually indicates a peak-to-peak airspeed excursion, a gain of half the reported value followed by a loss of the reported value. For an asymmetric microburst, these values can further vary. By better informing flight crews about the details of the alerts and the limitations of the sensor system, the inevitable "nuisance alerts" that will be issued will not hurt crew confidence in the system.

Acknowledgements

This work was supported by the MIT Lincoln Laboratory under contract BARR-10-119, and the Federal Aviation Administration and the National Aeronautics and Space Administration under grants NGL-22-009-640 and NAG-1-690. The authors would like to thank the respondents, Rick Brown, and United Airlines for assistance with the pilot opinion surveys. The authors would also like to thank Divya Chandra, Steven Bussolari, Ed Hahn, Amy Pritchett, ALPA, and especially the pilots who volunteered their time and wisdom, all of whom contributed to the success of the simulator experiments.

References

1. National Research Council, *Low Altitude Wind Shear and Its Hazard to Aviation*, National Academy Press, 1983.

2. Wolfson, M.M., "Characteristics of Microbursts in the Continental United States," *The Lincoln Laboratory Journal*, Vol. 1, No. 1, Spring 1988, pp. 49-74.

3. Chandra, Divya, *An Evaluation of Automation for Flight Path Management in Transport Category Aircraft*, MS Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, August, 1988.

4. Hart, S.G., and Staveland, L.E. (in press)., "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in P.A. Hancock and N. Meshkati (Eds.), *Human Mental Workload*, Elsevier Science Publishers, Amsterdam.

5. Targ, R., and Bowles, R.L., "Investigation of Airborne Lidar for Avoidance of Windshear Hazards," AIAA Paper 88-4658, Sept. 1988.

6. National Center for Atmospheric Research, "Terminal Doppler Weather Radar (TDWR): A Briefing Paper," July 1, 1988.

7. Sand, W., and Biter, C., "TDWR Display Experiences," AIAA Paper 89-0807, January 1989.

8. Stevenson, Lloyd, Draft Copy of "A PIREP-Based Analysis of the Candidate TDWR-Based Products and Services Evaluated at Stapelton International Airport During the Summer of 1988," Federal Aviation Administration Project Memorandum DOT-TSC-FA9E 1-89, May 1989.

9. Oseguera, Rosa M., and Bowles, Roland L., "A Simple Analytic 3-Dimensional Downburst Model Based on Boundary Layer Stagnation Flow," NASA TM-100632, NASA Langley Research Center, Hampton, VA, July, 1988.

10. Eilts, Michael D., "Estimation of Microburst Asymmetry with a Single Doppler Radar," Preprints, *Third International Conference on the Aviation Weather System*, January 30 - February 3, 1989, Anaheim, CA.

11. Elmore, Kimberly L., and Sand, Wayne R., "A Cursory Study of F-Factor Applied to Doppler Radar," Preprints, *Third International Conference on the Aviation Weather System*, January 30 - February 3, 1989, Anaheim, CA.

12. Bracalente, E.M., Britt, C.L., and Jones, W.R., "Airborne Doppler Radar Detection of Low Altitude Windshear," AIAA Paper 88-4657, Sept. 1988.

13. Adamson, H.P., "Airborne Passive Infrared System for the Advance Warning of Low-Level Windshear and Clear Air Turbulence: 1988 In-Service and Theoretical Work," AIAA Paper 88-4659, Sept 1988.

14. Federal Aviation Administration, *Windshear Training Aid*, 1987.

15. Hansman, R. J., and Wanke, C., "Cockpit Display of Hazardous Windshear Information," AIAA Paper 89-0808, January, 1989.

16. Frost, W., Chang, H. P., Elmore, K. L., and McCarthy, J., "Simulated Flight Through JAWS Wind Shear: In-Depth Analysis Results," AIAA Paper 84-0276, January, 1984.

17. Campbell, Steven D., and Isaminger, Mark A., "Using Features Aloft to Improve Timeliness of TDWR Hazard Warnings," Preprints, *Third International Conference on the Aviation Weather System*, January 30 - February 3, 1989, Anaheim, CA.

18. Psiaki, M. L., and Stengel, R. F., "Analysis of Aircraft Control Strategies for Microburst Encounter," *Journal of Guidance, Control, and Dynamics*, Vol. 8, No. 5, Sept-Oct 1985.

19. A'vila De Melo, Denise, *Analysis of Aircraft Performance During Lateral Maneuvering for Microburst Avoidance*, MS Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, September 1989.

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT OHIO UNIVERSITY 1989-1990

Robert W. Lilley
Department of Electrical and Computer Engineering
Ohio University
Athens, Ohio

SUMMARY OF RESEARCH

The Joint University Program (JUP) offers to students, faculty and professional staff the opportunity to contribute to the utility and safety of air transportation. At Ohio University, the program is structured to offer these opportunities to both graduate and undergraduate students. The program demands high-quality technical work, quarterly presentation and defense of the work, and publication at conferences and in journals. Under the general direction of the Principal Investigator, Joint University Program operations are led day-to-day by a selected Ph.D. candidate, who gains both technical and project leadership skills. During 1989-90, Frank van Graas, having achieved his Ph.D., has been gradually turning over the program to Michael Braasch. The annotated bibliography given below documents the technical activities and outputs for the year. Perhaps more important are the continuing careers of 1989-90 Joint University Program graduates:

James D. Waid - Graduated with the BSEE. He is currently with NASA Ames Research Center and will begin a Master's program at Stanford University in September 1990.

David S. McIntyre - Received the MSEE and is now with The Analytic Sciences Corporation (TASC), working with aircraft antenna systems.

Craig B. Parker - Completed the MSEE and is now with the Bendix-King Radio Company, a major avionics manufacturer.

Mark R. Kuhl - Received the MSEE and is now with Ashtech Inc., a GPS receiver manufacturer.

Paul A. Kline - Graduated with the BSEE with high honors, and remains with the Joint University Program as a graduate intern.

These JUP alumni carry both education and training to their new positions; they leave to the JUP their accomplishments while in school. The new students joining the program build on this plateau.

The JUP students have been regular (and very successful) entrants in the RTCA Jackson Award competition. Frank van Graas' receipt of the award for 1989 recognizes both his work and the program which made it possible. Messrs. Michael Braasch and David McIntyre won honors at the 2nd International ION Satellite Division meeting. Once again, this demonstrates the value of the JUP in preparing young engineers and scientists for effective careers.

During 1989-90, five topics received emphasis:

- A spectrum-efficient weather data uplink system was designed, constructed and flight-tested. Benefits from such systems can include an increase in safety by having real-time weather in convenient graphical form available to the pilot at all times. With data compression as shown by Craig Parker, existing communication or navigation systems may be used to uplink (broadcast) the data.

- Integrated GPS/INS study continued, utilizing the Redundant Strapdown IMU on loan from NASA LaRC. Hybrid and integrated systems will become the norm on all classes of aircraft, and any added knowledge of multi-sensor navigation and guidance systems will prove helpful. GPS and INS are an ideal hybrid pair, as are GPS and Loran-C and other such partners. Increased navigation availability and integrity result.

- The Ridge Regression theory was refined and applied to air navigation scenarios. The extent to which this technique can improve navigation accuracy with only a modest computing load and its comparison with Kalman filters remain active areas of study. If initial promising results are validated, the application of this technique could result in cost savings due to the need for fewer ground emitters. Alternatively, higher availability and integrity numbers could result due to the relaxation permitted in emitter geometries relative to the aircraft.

- System Identification theory was applied to GPS data, to point the way to better understanding of the effects of Selective Availability (S/A) on civilian users of this navigation system. Even differential GPS configuration designers need to consider S/A characteristics in their work.

- JUP meetings provide a forum for reporting on other scientific work of interest to the group. Analysis of thought-related (electroencephalographic) signals for application to control of computer systems could have significance in aid to paraplegics, or for hands-off system control in industrial or air-traffic control areas.

The FAA/NASA Joint University Program is successful and mature because of its stability of sponsor support. The predictable nature of the JUP attracts quality students at all levels, with the assurance that their academic and research activities can be completed without interruption. These students then produce, in partnership with faculty and staff mentors, quality technical products and build the bases for successful careers supporting air transportation.

ANNOTATED BIBLIOGRAPHY OF 1989-90 PUBLICATIONS

1. McFarland, R. H., and Parker, C. B.: Weather Data Dissemination to Aircraft. Proceedings of the AIAA 27th Aerospace Sciences Meeting, AIAA Paper 89-0809, Reno, NV, January 9-12, 1989.

Documentation exists that shows weather to be responsible for approximately 40 percent of all general aviation accidents with fatalities. Weather data products on the ground are becoming more sophisticated and greater in number. Although many of these data are critical to aircraft safety, they currently must be transmitted verbally to aircraft. This process is labor intensive and provides a low rate of information transfer. Consequently, the pilot is often forced to make life-critical decisions based on incomplete and outdated information.

Automated transmission of weather data from the ground to aircraft can provide aircrew with accurate data in near real time. The current National Airspace System Plan calls for such an uplink capability to be provided by the Mode S beacon system data link. Although this system has a very advanced data link capability, it will not be capable of providing adequate weather data to all airspace users in its planned configuration. This paper delineates some of the important weather data uplink system requirements, and describes a system which is capable of meeting these requirements. The proposed system utilizes a run-length coding technique for image data compression and hybrid phase and amplitude modulation techniques for the simultaneous transmission of both voice and weather data on existing aeronautical Very High Frequency (VHF) voice communication channels.

2. Parker, C. B.: A Technique for the Automated Dissemination of Weather Data to Aircraft. Master's Thesis, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, June 1989.

The issue of the provision of weather data products to the aircraft cockpit is addressed. The availability of these data in the cockpit is necessary to allow the pilot to make effective go/no-go decisions when flying in threatening weather conditions. Currently, except for the limited text capabilities of the ARINC Communication Addressing and Reporting System (ACARS), weather data are disseminated by voice communication between aircraft and air traffic control (ATC) personnel. This process severely limits the ability of the pilot and aircrew to obtain timely and accurate weather data.

The transmission of weather data products to the aircraft by data uplink is widely recognized as the solution to the weather data dissemination problem. An analysis is presented in this paper which shows that existing VHF communication systems can simultaneously transmit both analog voice and digital weather data. This hybrid modulation can be accomplished without unacceptable degradation to either communication mode.

An experimental system for the uplink of graphical weather data products is described. This system was developed in order to provide for the evaluation of graphic weather image compression algorithms and cockpit weather displays. Typical compression ratios from 6:1 to 9:1 have been demonstrated. These images have been successfully transmitted in 8 to 12 seconds at 2,400 bits per second.

Finally, it is proposed that the system developed in this thesis be used for the simultaneous transmission of both recorded voice and digital weather data from existing Automatic Terminal Information Service (ATIS) VHF transmitters to the aircraft.

3. Parker, C. B.: 2400 Bit/Second Modem for Audio Channels. Technical Memorandum OU/AEC 30-89TM NASA TRI-U/122, Avionics Engineering Center, Department of ECE, Ohio University, Athens, OH, July 1989.

A transmit and receiver modem pair is described which is capable of transmitting digital data at a data rate of 2400 bits per second over an audio channel.

4. McIntyre, D. S.: Integrated GPS/INS Attitude and Heading Determination. Proceedings of the 2nd International Satellite Division Meeting of the ION, Student Session, Colorado Springs, CO, September 27-29, 1989.

The phase margin requirements of the effective GPS receiver data rate to the guidance loop of the flight control system are discussed. An in-flight attitude and heading algorithm implementing the synergistic benefits of an integrated GPS/INS system is also presented. An optimum approach to the derivation of position and velocity data to obtain aircraft orientation is outlined. In closure, an application of the GPS/INS algorithm to the attitude and heading determination of hypersonic aircraft is presented.

5. McIntyre, D. S.: GPS Effective Data Rate Characterization with Application to Integrated GPS/INS Attitude and Heading Determination. Master's Thesis, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, June 1989.

The integration of the NAVSTAR Global Positioning System (GPS) with the Inertial Navigation System (INS) offers many benefits. INS drift errors may be compensated by the long-term stability of GPS. Conversely, the short-term stability of the INS can be used to correct or substitute for faulty GPS data. Dissimilar redundancy is an additional advantage.

The optimization of the effective GPS data rate is essential for the proper operation of an integrated GPS/INS. This paper develops a transfer function for the GPS receiver. A receiver simulation was developed and tested to determine if the phase response of the transfer function was acceptable for an accurate transmission of attitude and heading information to the flight control system. After the results were interpreted, an integrated GPS/INS algorithm was presented for providing accurate attitude and heading information to a High Speed Civil Transport (HSCT) aircraft guidance control system.

6. van Graas, F., Waid, J. D., and Kline, P. K.: Flight-Test Results for a Prototype Hybrid GPS/LORAN Receiver. Proceedings of the 18th Annual Technical Symposium of the WGA, Hyannis, MA, Oct. 29 - Nov. 1, 1989.

Earth-referenced navigation based on the NAVSTAR Global Positioning System (GPS) and the Long Range Navigation System, LORAN-C, has the potential to satisfy the requirements for a sole means of navigation for the conterminous United States. This paper presents the

design considerations and architecture of a prototype hybrid GPS/LORAN receiver. The receiver is installed in a research aircraft to evaluate the in-flight performance of hybrid GPS/LORAN. The flight-test data is referenced to a differential GPS truth trajectory. Initial test results demonstrate hybrid GPS/LORAN accuracies consistent with current requirements for en route and terminal navigation, and non-precision approaches.

7. Vicksell, F. B., Goddard, R. B., Enge, P. K., and van Graas, F.: Analysis of LORAN-C/GPS Interoperability for Air Navigation. Proceedings of the 18th Annual Technical Symposium of the WGA, Hyannis, MA, Oct. 29 - Nov. 1, 1989.

Computer runs show that it is likely that in the National Airspace System a hybrid of LORAN-C and a 24-satellite Global Positioning System can meet aviation sole means requirements for availability and accuracy, and perhaps integrity as well. A hybrid GPS/LORAN system reduces unavailability by a factor of 1000 compared to GPS alone.

The GPS and the LORAN-C system signals are well suited for combination in a hybrid fix algorithm. Virtual synchronization of GPS and LORAN clocks can be achieved by inclusion of each LORAN transmitter's offset from Universal Time within the LORAN signal. GPS failure rates and distributions of GPS Selective Availability are not yet known, nor are the characteristics of rare high levels of atmospheric noise affecting LORAN measurements known; reasonable estimates were used. An integrity requirement specification would include both a maximum miss rate and a maximum alarm rate along with the radial protection limit. For integrity checking, both the maximum separation and least squares residuals techniques were examined.

8. Kelly, R. J., van Graas, F., and Kuhl, M. R.: Improved Effectiveness of GPS RAIM through Ridge Regression Signal Processing. Proceedings of the 2nd International Satellite Division Meeting of the ION, Colorado Springs, CO, September 27-29, 1989.

A new measurement processing technique is presented which significantly improves the effectiveness of GPS Receiver Autonomous Integrity Monitoring (RAIM). RAIM is a software-based algorithm implemented in a GPS receiver to detect satellite signal failures. The detection of a satellite failure is based on the consistency of a redundant set of pseudorange measurements. GPS positioning requires four satellites to solve for three-dimensional position and clock offset from GPS time. When five satellites are in view, five different sub-solutions can be calculated, each omitting one of the five satellites. If no satellite failures are present, the five sub-solutions will be close to one another. However, if one of the satellites has failed, the sub-solutions will be scattered. The integrity alarm is triggered based on comparisons of the sub-solutions. Unfortunately, a poor distribution of the satellites also causes the sub-solutions to be scattered. Poor geometry, usually expressed in terms of the horizontal dilution of precision (HDOP), frequently exists for some sub-solutions, thus limiting the availability of RAIM. As the HDOP increases, position errors are correspondingly inflated.

In 1970, Hoerl and Kennard developed Ridge Regression to combat near collinearity when it arises in the predictor matrix of a linear regression model. Since a nearly collinear predictor matrix is directly related to a large HDOP, Ridge Regression processing can be applied to reduce the effect of the HDOP, thus permitting GPS to work. It is also shown that the effects of near collinearity cannot be reduced by choosing the estimator error covariance matrix Q_0

such that the innovations of the Kalman filter are minimized. However, a methodology for selecting Q_0 is presented which incorporates Ridge Regression into the Kalman filter. Finally, a computer simulation is implemented to illustrate the improved availability of GPS RAIM by using Ridge Regression processing to reduce those sub-solution errors due to a high HDOP.

9. Braasch, M. S.: A Signal Model for the NAVSTAR Global Positioning System. Proceedings of the 2nd International Satellite Division Meeting of the ION, Student Session, Colorado Springs, CO, September 27-29, 1989.

As the development of the NAVSTAR Global Positioning System (GPS) continues, there will increasingly be the need for a software-centered signal model. This model must accurately generate the observed pseudorange which would typically be encountered. The observed pseudorange varies from the true geometric (slant) range due to range measurement errors. Errors in the range measurements stem from a variety of hardware and environment factors. These errors are classified as either deterministic or random and, where appropriate, their models are summarized. Of particular interest is the model for Selective Availability (S/A) which was derived from actual GPS data. The procedure for the determination of this model, known as System Identification Theory, is briefly outlined. The synthesis of these error sources into the final signal model is given along with simulation results.

RIDGE REGRESSION SIGNAL PROCESSING

Mark R. Kuhl
Ohio University
Athens, Ohio

SUMMARY

The introduction of the Global Positioning System (GPS) into the National Airspace System (NAS) necessitates the development of Receiver Autonomous Integrity Monitoring (RAIM) techniques as identified by RTCA Special Committee 159. In addition, it is anticipated that future multisensor positioning systems will also utilize RAIM schemes to guarantee a high level of integrity, as evidenced by current FAA programs involving hybrid positioning systems such as GPS/LORAN-C.

In order to guarantee a certain level of integrity, a thorough understanding of modern estimation techniques applied to navigational problems is required. In this paper, the extended Kalman filter (EKF) is derived and analyzed under poor geometry conditions. It was found that the performance of the EKF is difficult to predict, since the EKF is designed for a Gaussian environment. A novel approach is implemented which incorporates ridge regression to explain the behavior of an EKF in the presence of dynamics under poor geometry conditions.

The basic principles of ridge regression theory are presented, followed by the derivation of a linearized recursive ridge estimator. Computer simulations are performed to confirm the underlying theory and to provide a comparative analysis of the EKF and the recursive ridge estimator.

BACKGROUND OF RIDGE REGRESSION

Ridge regression signal processing is introduced in order to counter the effects of poor geometry on position estimation. Robert Kelly first applied ridge regression to navigational problems in 1988 (ref. 1). Many other papers have followed since then explaining the further developments of ridge regression (refs. 1-6). The basic idea behind ridge regression is presented below.

For navigation, the appropriate error criterion is the mean square error (MSE). It expresses the deviation of the vehicle with respect to its intended path. The MSE consists of the sum of two error components; MSE = Variance + Bias^2. For ordinary least squares (OLS) estimation, the position solution is unbiased; therefore, the MSE is only made up of the variance term. (Note that the Kalman filter is a recursive form of the OLS estimator.) Under poor geometry conditions, the variance of the position errors inflates significantly, giving an inaccurate position estimate. The ridge estimator takes advantage of an extra degree of freedom in the MSE which is not used by the OLS estimator - the bias term. In effect, small biases induced by the ridge estimator decrease the variance term such that the overall MSE is

smaller than the MSE obtained from an unbiased estimator, as illustrated in figure 1.

The linear model for a system with an unknown n x 1 measurement bias vector, $\underline{\delta B}$, and a measurement noise vector, $\underline{\epsilon}$, is given by

$$\underline{Y} = H\underline{\beta} + \underline{\delta B} + \underline{\epsilon} \tag{1}$$

where $\underline{Y}$ is the n x 1 range measurement vector, $\underline{\beta}$ is the p x 1 unknown system state vector (or parameter vector), and H is the n x p predictor (or design) matrix which relates the range measurement vector to the system state vector. Also, the measurement noise is uncorrelated; $\text{cov}(\underline{\epsilon} + \underline{\delta B}) = [\underline{\epsilon}\underline{\epsilon}^T] = \sigma^2 I$, where I is the n x n identity matrix.

The OLS estimate of equation 1 is

$$\hat{\underline{\beta}}_{OLS} = (H^T H)^{-1} H^T \underline{Y} \tag{2}$$

The corresponding ridge estimate of equation 1 is

$$\hat{\underline{\beta}}_R = (H^T H + P_R)^{-1} H^T \underline{Y} \tag{3}$$

where P_R is the ridge parameter matrix (which is formed from the chosen ridge biasing parameters, k_i). When P_R consists of zeros, the ridge estimator reduces to the OLS estimator. Adding a non-zero ridge parameter matrix to $H^T H$ purposely upsets the balance between the first and second moment components of $\hat{\underline{\beta}}$, thereby introducing a bias.

For a complete discussion on the comparative analysis of the ridge and OLS estimators, see reference 6.

THE EXTENDED KALMAN FILTER (EKF)

From reference 6, figure 2 summarizes how the normal Kalman equations relate to each other to form the Kalman filter. Since the equations that relate the measurements to the state vector are usually nonlinear (i.e. the H matrix is nonlinear), an extended Kalman filter (EKF) is needed. Therefore, a linearization procedure is performed when deriving the Kalman filter equations, see reference 7.

After the linearization procedure, the recursive update equation for the extended Kalman filter is given by

$$\hat{\underline{\beta}}_k = \hat{\underline{\beta}}_{k/k-1} + (P^{-1}_{k/k-1} + H_k^T R^{-1} H_k)^{-1} (H_k^T R^{-1} \delta \underline{Y}_k) \tag{4}$$

where $\delta \underline{Y}_k$ is the difference between the actual measurements and the predicted measurements (the innovations) as shown below

$$\delta \underline{Y}_k = \underline{Y}_k - g_k(\hat{\underline{\beta}}_{k/k-1}) \tag{5}$$

where g_k is a nonlinear vector-valued function which relates the system state to the measurements.

An optimum unbiased estimator arises when both the model and the estimator match the process which generates the data. Kalman filter optimization techniques include the selection of the initial error covariance matrix, P_0, and the system covariance matrix, Q, based upon past experience or by adaptively tuning the filter until its innovations (residuals) become white (i.e. zero mean and random). However, the selection of the proper Q matrix is usually not a very easy task (ref. 8).

For example, the system covariance matrix Q is often set artificially high such that the Kalman filter can track the vehicle when it encounters dynamics such as turning-induced accelerations. Therefore, the Kalman filter allows more noise in the solution during periods of low dynamics. Although some methods exist for selecting Q adaptively (ref. 9), these are stochastic in nature.

Another problem arises when the Kalman filter is subjected to a poor geometry condition. In the case of inaccurate P_0 and Q matrices (a mismatched filter), the filter may become biased. The performance of a biased Kalman filter is not readily understood, as the Kalman filter is optimal and defined in a Gaussian environment only. Therefore, the performance of a Kalman filter for a deterministic maneuver in a poor geometry condition cannot be predicted from the regular Kalman filter equations. Furthermore, poor geometry effects cannot be minimized by observing the innovations in the Kalman filter's update equation (see equation 4), because these effects only appear in the estimator $\hat{\underline{\beta}}$.

In applying the Kalman filter to navigation, the MSE is the appropriate error criterion to be minimized. The MSE of a mismatched Kalman filter is not necessarily the smallest obtainable. Recall that the MSE is the sum of two components: the variance term and bias term squared. Since the Kalman filter is developed from the OLS estimator, it is inherently an unbiased estimator. On the other hand, a biased estimator, such as the ridge recursive filter presented in the next section, is purposely not matched to the process that generates the data in order to achieve a smaller MSE.

THE RIDGE/EKF FILTER

Extending the work presented in reference 5, the linearized state update equation for the extended Ridge/EKF is obtained as

$$\hat{\beta}_{Rk} = \hat{\beta}_{Rk/k-1} + (P^{-1}_{k/k-1} + P_{Rk} + H_k^T R^{-1} H_k)^{-1}(H_k^T R^{-1} \delta \underline{Y}_k) \qquad (6)$$

where P_{Rk} has been added to the $(P^{-1}_{k/k-1} + H_k^T R^{-1} H_k)$ term in order to reduce the effects of poor geometry. Also, the update equation for the error covariance matrix is given by

$$P_k = (P^{-1}_{k/k-1} + P_{Rk} + H_k^T R^{-1} H_k)^{-1}. \qquad (7)$$

Two cases may now be defined in which ridge regression can explain the behavior of the Kalman filter. First, in the absence of system dynamics, the ridge parameter matrix, P_{Rk} is functionally equivalent to the inverse of the state error covariance matrix, $P^{-1}_{k/k-1}$. Therefore, the EKF has similar convergence properties when the model incorrectly represents the system and $P_{k/k-1}$ is small. (See reference 6 for a discussion on the convergence properties of estimators.) Second, if dynamics exist, P_{Rk} is related to both $P^{-1}_{k/k-1}$ and the system error

covariance matrix, Q, given by the following relation

$$(\Phi P_k \Phi^T + Q)^{-1} \sim (\Phi P_k \Phi^T + Q)^{-1} + P_{Rk} \quad (8)$$

where, in this case, the ~ symbol means "relating to". Again, the EKF has similar convergence properties when the model incorrectly represents the system and $(\Phi P_k \Phi^T + Q)^{-1}$ is small.

Usually P_k is small, which means that the $P^{-1}_{k/k-1}$ term in equation 6 is large; therefore, there is no geometry problem. Note that P_k varies as the Kalman filter is updated, but normally a constant Q is added which "limits" it (i.e. puts uncertainty in the model). As seen in the above equation, P_k will be large when Q is large. Q is chosen large for dynamic situations (i.e. turn-induced accelerations). When a poor geometry condition exists in addition to the dynamic situation, P_{Rk} may be added to counteract the large Q. Therefore, a proper P_{Rk} can be chosen to incorporate ridge regression into the Kalman filter.

The key idea in developing the ridge recursive filter is the following: Each step in the recursive process is viewed as a new prior linear model wherein the last estimate $\hat{\beta}_{Rk/k-1}$ is the prior equation for the next iteration. The ridge solution is recomputed at each step using selection rules, described in reference 6, to determine a proper P_{Rk}.

SIMULATION RESULTS

This section provides a simulation to evaluate the behavior of an extended Kalman filter (EKF) when dynamics exist and a poor geometry condition occurs. Initial results show how ridge regression can explain the behavior of a mismatched Kalman filter.

The EKF is implemented in the following simulation scenario: An aircraft is moving along a constant velocity flight path making range measurements to two Distance Measuring Equipment (DME) stations. At first (t = 0 seconds), the station locations constitute a good geometry situation with respect to the aircraft position. At t = 50 seconds, the aircraft "enters a poor geometry condition". This happens when the aircraft switches from making range measurements to DME stations 1 and 2 to making range measurements to DME stations 1 and 3. After some time in this poor geometry condition, the aircraft turns 30° to the right (t = 150 seconds) and continues in this situation until the end of the simulation run (t = 300 seconds). The numbers for the scenario set-up are given below:

DME station 1 location, $[X_1\ Y_1]$ = [10.0e05 0] feet
DME station 2 location, $[X_2\ Y_2]$ = [85.0e05 75.0e05] feet
DME station 3 location, $[X_3\ Y_3]$ = [160.0e05 0] feet

initial aircraft location, $[X_{I_R}\ Y_{I_R}]$ = [85.0e05 -1.0e05] feet
initial aircraft velocity, $[V_x\ V_y]$ = [100 0] feet/sec
aircraft velocity after turn, $[V_x\ V_y]$ = [86.6 50] feet/sec
measurement noise, σ = 30 feet
measurement bias, $\underline{\delta B}$ = [0 0] feet
aircraft initial position offset, [δX δY] = [100 100] feet

The aircraft makes 16 measurements to each DME station per second and updates its position estimate once every 16^{th} of a second (i.e. recursive estimation).

The initialization of the tuned Kalman filter is as follows: The initial estimated system state vector and its associated error covariance matrix are

$$\hat{\beta}_0 = \begin{bmatrix} \hat{X} \\ \hat{Y} \\ \hat{V}_x \\ \hat{V}_y \end{bmatrix} = \begin{bmatrix} 85.0e05 - 100 \\ -1.0e05 - 100 \\ 0 \\ 0 \end{bmatrix} \qquad P_0 = \begin{bmatrix} 100^2 & 0 & 0 & 0 \\ 0 & 100^2 & 0 & 0 \\ 0 & 0 & 100^2 & 0 \\ 0 & 0 & 0 & 100^2 \end{bmatrix}$$

The system process covariance matrix is

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{bmatrix}$$

And the measurement noise covariance matrix is

$$R = \begin{bmatrix} 30^2 & 0 \\ 0 & 30^2 \end{bmatrix}$$

The results of the tuned Kalman filter for the above scenario are portrayed in figure 3 where the estimated flight path (each estimate is represented by an "o") is compared to the true flight path (represented by a straight line). Note that there is some lag in the Kalman filter estimates when the aircraft starts to turn. The square root of the MSE is 134 feet. This result shows that the Kalman filter can be properly tuned to handle dynamics when conditions of poor geometry arise.

Normally, the Q matrix in the Kalman filter is set artificially high such that the filter can follow a vehicle through periods when dynamics exist. In order to see how ridge regression can reduce the square root of the MSE, implement the same simulation set-up provided above with a much higher Q matrix.

$$Q = \begin{bmatrix} 500 & 0 & 0 & 0 \\ 0 & 500 & 0 & 0 \\ 0 & 0 & 500 & 0 \\ 0 & 0 & 0 & 500 \end{bmatrix}$$

The estimated and true aircraft flight paths are shown in figure 4. The square root of the MSE becomes 304 feet.

Now, add a proper ridge parameter matrix, P_g, to counteract the effects of poor geometry. In effect, adding P_g is similar to making Q smaller. Figure 5 depicts the resulting estimated and true aircraft flight paths. The square root of the MSE reduces to 142 feet. This shows that a properly selected P_g matrix can give results that are comparable to the tuned Kalman filter described above. Efforts are continuing to address the performance of the extended Ridge/EKF in detail.

CONCLUSIONS

Ridge regression signal processing is introduced to explain the behavior of a mismatched Kalman filter. This paper provides an initial comparison of the extended Kalman filter (EKF) and the recursive ridge filter (the Ridge/EKF). It is shown that a proper ridge parameter matrix P_g, may be chosen to counteract poor geometry conditions that may arise when estimating a position.

REFERENCES

1. Kelly, R. J.: Reducing Geometric Dilution of Precision Using Ridge Regression Signal Processing. Plans '88, Orlando, FL, November 29 through December 2, 1988.

2. Kelly, R. J.: OMEGA Accuracy Improvement Using Ridge Regression. OMEGA: International and Economic, Proceedings of the Fourteenth Annual Meeting, Long Beach, CA, 1989.

3. Kelly, R. J.: Global Positioning System Accuracy Improvement Using Ridge Regression. AIAA Guidance, Navigation and Control Conference, Boston, MA, August 14-16, 1989.

4. Kelly, R. J., van Graas, F., and Kuhl, M. R.: Improved Effectiveness of GPS RAIM through Ridge Regression Signal Processing. Proceedings of the 2nd International Satellite Division Meeting of the ION, Colorado Springs, CO, September 27-29, 1989.

5. Kelly, R. J.: Reducing Geometric Dilution of Precision Using Ridge Regression. IEEE Transactions on Aerospace and Electronic Systems, Vol. 26, No. 1, January 1990.

6. Kuhl, M. R.: Ridge Regression Signal Processing Applied to Multisensor Position Fixing. Master's Thesis, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, June 1990.

7. Ewing, P. L.: The Design and Implementation of Tracking and Filtering Algorithms for an Aircraft Beacon Collision Warning System. Master's Thesis, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, March 1989.

8. Brown, R. G.: Introduction to Random Signal Analysis and Kalman Filtering. John Wiley and Sons, New York, 1983.

9. Sorenson, H. W.: Kalman Filtering: Theory and Applications. IEEE Press, New York, 1985.

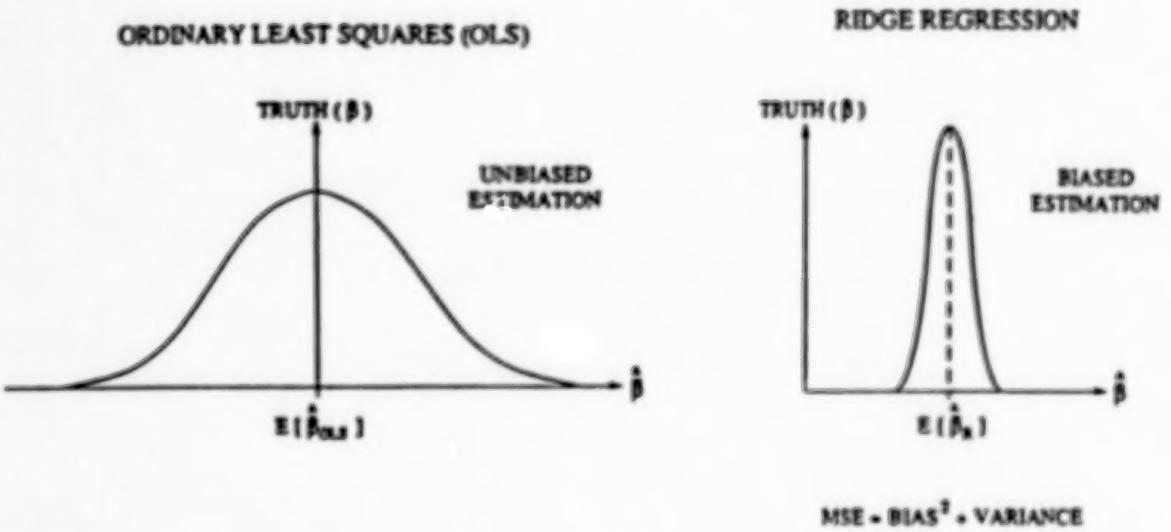


Figure 1. Comparison of unbiased and biased solution distributions and definition of the Mean Squared Error (MSE).

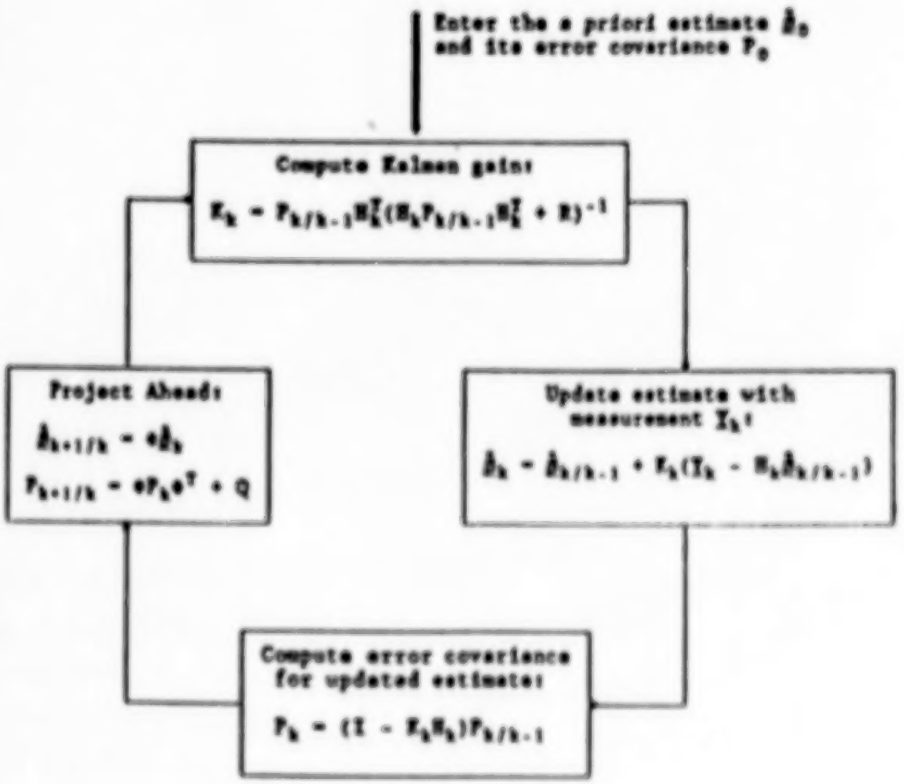

Enter the a priori estimate $\hat{x}_0$ and its error covariance P_0
Compute Kalman gain:
$K_k = P_{k/k-1}H_k^T(H_kP_{k/k-1}H_k^T + R)^{-1}$
Project Ahead:
$\hat{x}_{k+1/k} = \Phi\hat{x}_k$
$P_{k+1/k} = \Phi P_k\Phi^T + Q$
Update estimate with measurement Y_k:
$\hat{x}_k = \hat{x}_{k/k-1} + K_k(Y_k - H_k\hat{x}_{k/k-1})$
Compute error covariance for updated estimate:
$P_k = (I - K_kH_k)P_{k/k-1}$

Figure 2. Kalman filter equations for a linear system.

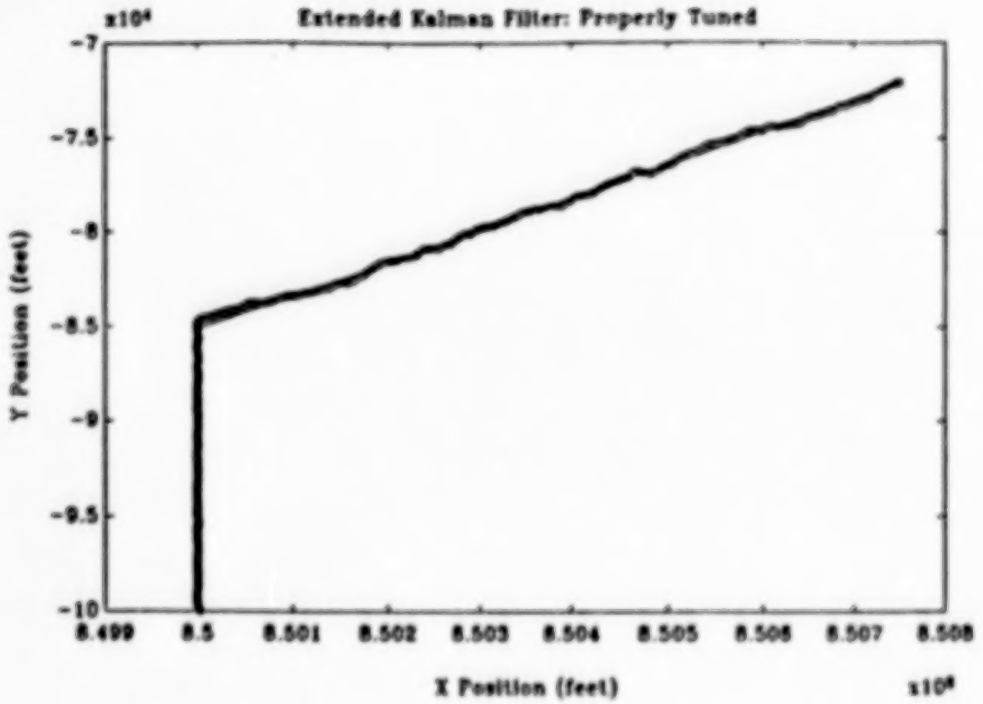


Figure 3. Simulation results for a properly tuned extended Kalman filter (small system covariance).

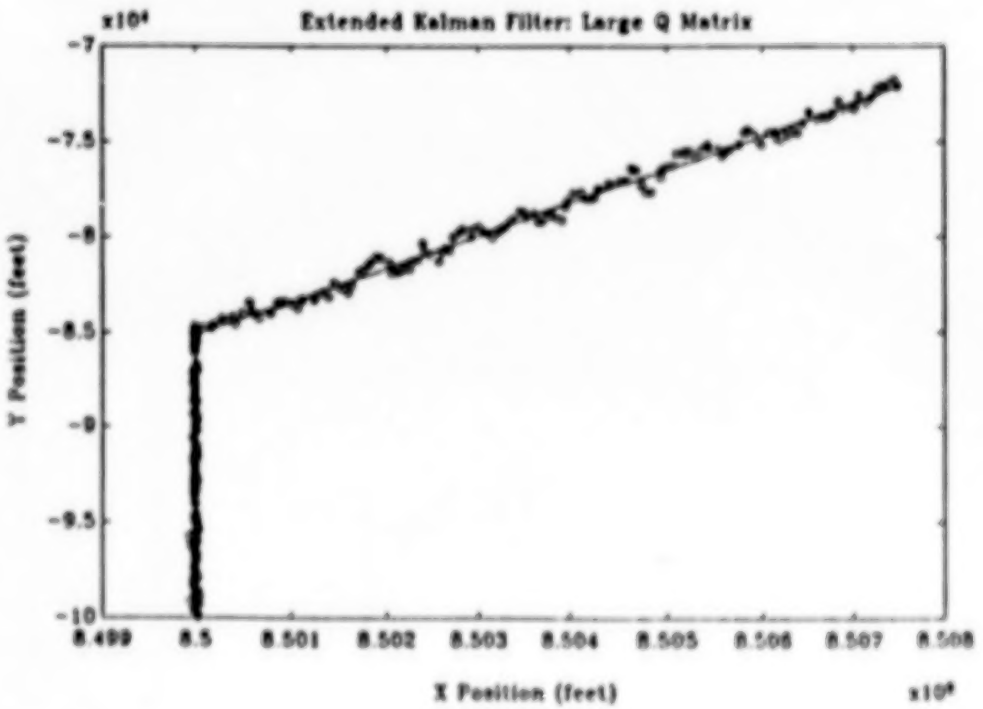


Figure 4. Simulation results for an extended Kalman filter allowing for dynamics (large system covariance).

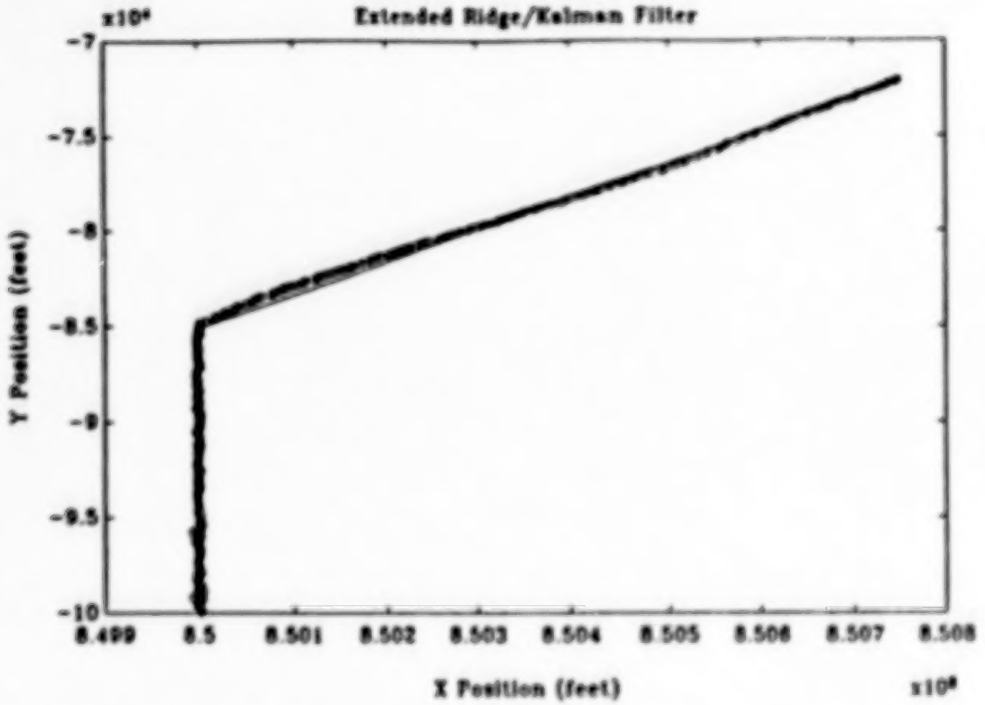


Figure 5. Simulation results for a Ridge/EKF filter.

MODELING SELECTIVE AVAILABILITY OF THE NAVSTAR GLOBAL POSITIONING SYSTEM

Michael Braasch
Ohio University
Athens, Ohio

SUMMARY

As the development of the NAVSTAR Global Positioning System (GPS) continues, there will increasingly be the need for a software-centered signal model. This model must accurately generate the observed pseudorange which would typically be encountered. The observed pseudorange varies from the true geometric (slant) range due to range measurement errors. Errors in range measurements stem from a variety of hardware and environment factors. These errors are classified as either deterministic or random and, where appropriate, their models are summarized. Of particular interest is the model for Selective Availability (S/A) which was derived from actual GPS data. The procedure for the determination of this model, known as System Identification Theory, is briefly outlined. The synthesis of these error sources into the final signal model is given along with simulation results.

INTRODUCTION

The NAVSTAR Global Positioning System (GPS) is a satellite radionavigation system designed to provide worldwide position, velocity, and time information. The use of GPS may be extended to land, air, sea, and space. For civil air use however, system performance must be proven before regulatory agencies will sanction its use. Strict requirements are imposed and documentation must be produced which demonstrates the system's ability to meet these requirements.

Of these strict requirements, the most difficult to satisfy are low failure probability and high integrity. The role of the regulatory agencies is also seen in the certification of airborne GPS equipment. Before a GPS receiver is certified, documentation must be produced which details the GPS performance. Since the cost of flight testing is high, the majority of this documentation is produced through simulations. One input to these simulations is a model of the navigation system.

This paper presents a signal model which may be used for such purposes. Error source algorithms are generated into one model which computes the observed pseudorange for each satellite in view and determines the perturbed user position. This pseudorange model may be used in the performance analysis of GPS, differential GPS (DGPS), and receiver autonomous integrity monitoring (RAIM) (ref. 1). In addition, the model may be used to determine the required update rate for DGPS (ref. 2).

GPS ERROR SOURCES AND MODELS

The pseudorange observation equation is given by:

$$P_r = P_s + t_{rcvr} - t_{sv} + d_{tropo} + d_{iono} + URE + d_{mp} + d_{HW} + d_{meas} + noise + S/A$$

where:

P_r The observed pseudorange as determined by the receiver, which is the difference between the time the signal was transmitted by the satellite and the time the signal arrived at the receiver, multiplied by the speed of light.

P_s The true geometric (slant) range between the satellite and the receiver.

t_{rcvr} Receiver clock offset from GPS time, as estimated from observed pseudoranges.

t_{sv} Satellite clock offset from GPS time, as decoded from the satellite navigation data.

d_{tropo} Propagation delay caused by the troposphere.

d_{iono} Propagation delay caused by the ionosphere.

URE User range error.

d_{mp} Specular multipath errors.

d_{HW} Receiver hardware delay.

d_{meas} Receiver measurement bias errors.

noise Receiver measurement noise, clock noise, and diffuse multipath.

S/A Selective Availability.

Both the receiver clock offset and hardware bias are essentially identical for each satellite being tracked. As a result, these biases will appear as a clock bias in the user navigation solution and hardly affect the position solution.

The error models for tropospheric delay, ionospheric delay, and URE have been described in detail in the literature and require no further comment. It is the random errors which currently pose the greatest challenge.

Receiver noise, satellite clock noise, and Selective Availability are all stochastic processes and therefore cannot be predicted exactly. While the combination of the receiver and satellite clock noise is basically a white noise process with small variance, S/A is not. Moreover, the exact model which produces S/A is classified for security reasons. Since the actual model which produces S/A is classified, real GPS data from an operational satellite are needed. Once data have been collected which contain S/A, a more rigorous analysis may be performed. In addition, once S/A waveforms have been derived from the data, a technique is required which

determines the model which best fits the process which produced it. This technique is known as System Identification Theory.

SYSTEM IDENTIFICATION THEORY

Following reference 3, the essence of System Identification Theory will be described. The main concept is the idea that the data to be identified can be considered the output of a linear filter whose input is white noise. In general, a process may be described by an Auto-Regressive Moving Average (ARMA) model of a given order.

Collected data are passed through a filter whose coefficients are adjusted until its output is white noise (with variance σ_r^2). The final filter coefficients can then be used to find the inverse of the signal generator filter. Data which are statistically equivalent to the collected data may then be generated by passing white noise (with variance σ_r^2) through the inverse filter. This process is depicted in figure 1.

In summary, the steps which lead to the S/A model are:

1) Obtain S/A sample through data collection.

2) Determine a filter whose output is white noise when the S/A data sample is passed through it. Calculate the variance of the white noise output.

3) Construct the inverse of the filter.

4) Statistically equivalent S/A data may then be obtained by passing white noise (with variance as calculated in step 2) through the filter obtained in step 3.

In addition to the determination of the S/A model, System Identification Theory also offers a significant compression of the S/A data. Once the inverse model is determined, only the model coefficients and the variance of the residuals need to be stored. Based on these few parameters, virtually an infinite amount of statistically equivalent S/A data can be generated.

GPS DATA COLLECTION

As outlined in the above section, the first step in determining a model for S/A is to collect a sample of S/A. Data were collected from the first operational Block II satellite (SV14). Block II satellites are the first satellites capable of introducing S/A into the GPS signal. The satellite was tracked since day 135. On days 189, 190, and 191, error data which are likely to be caused by S/A were collected. As of the time of this writing, the S/A data presented in this paper have not been confirmed by JPO. However, this does not present a major problem as will be discussed later.

Figures 2 and 3 show plots of SV14 pseudorange residuals. Figure 2 shows the pseudorange residuals with S/A turned off. The error is composed mainly of receiver noise. Figure 3 shows what appears to be two different models of S/A.

DETERMINATION OF S/A MODELS

As outlined previously, once a data sample has been collected the next step in the identification process is to determine a proper filter. The proper filter has the following characteristics:

1) Its output is white noise when the data are passed through it.
2) It has a minimum number of coefficients.

Once this filter has been determined, the variance (σ_r^2) of the white noise output is calculated. The S/A generator is then composed of white noise (with variance σ_r^2) input to the inverse of the above described filter.

This procedure was carried out for the four S/A samples collected on days 189, 190, and 191. It is important to note that the data sets analyzed were unfiltered. This was done for two reasons. One was to avoid assumptions imposed on the S/A waveforms by the low-pass filter. The second was to more accurately reproduce the SV14 residuals. The generator generates waveforms which are statistically equivalent to the combination of the input S/A and noise.

The determination of the S/A model is illustrated for the data collected on day 190 (GPS time of the week: 22 - 24 hours). The model chosen was an ARMA model with three poles and two zeros, ARMA(3,2). In order to validate this model, the autocorrelation function of the output is calculated to check for whiteness (figure 4). Since it has a spike at lag 0 and its sidelobes are contained in the 95% confidence limits, it may be considered white noise. An additional test is to plot the power spectrum of the output. White noise is characterized by a flat spectrum. Again, the conclusion is verified (figure 5).

Now that the optimum filter has been determined, the waveform generator results from taking the inverse of the optimum filter. The resulting filter is given by:

$$H(z) = \frac{1 - 1.5047z^{-1} + 0.6428z^{-2}}{1 - 2.0833z^{-1} + 1.2164z^{-2} - 0.1302z^{-3}}$$

As a further check, data are generated and plotted with the original data (figures 6 and 7). Note: the generated data do not look exactly like the original because these data are statistically equivalent. The power spectrum of the collected data and the generated data show the spectral similarity (figure 8).

Table 1 gives the models which result from the identification of the four S/A samples. It is interesting to note that the models for day 189 (166 - 168) and day 190 (22 - 24) are very similar. The same is true for day 190 (0 - 2) and day 191 (24 - 26). It is probable that in

fact two models were used to create the four S/A waveforms.

This brings up an interesting point. As was mentioned in the Data Collection section, the collected waveforms have not been confirmed by JPO as S/A. However, if actual S/A turns out to be different than what has been determined, it is a simple matter to go through the ID process again and change the filter coefficients in the model.

SIGNAL MODEL BLOCK DIAGRAM AND SIMULATION RESULTS

The block diagram of the signal model is given in figure 9. The GPS ephemerides, the true user state, and the error sources combine in a truth model. The truth model outputs satellite positions and perturbed pseudoranges to the navigation solution block. The true user state is then subtracted from the calculated user state resulting in the user state error.

Table 1

MODELS IDENTIFIED FROM THE FOUR S/A SAMPLES

Day (GPS time of week in hours)	Model	σ_r^2
189 (166-168)	$H(z) = \frac{1 - 1.6814z^{-1} + 0.8082z^{-2}}{1 - 2.3161z^{-1} + 1.6630z^{-2} - 0.3450z^{-3}}$	20.6
190 (0-2)	$H(z) = \frac{1 - 1.5890z^{-1} + 0.6744z^{-2}}{1 - 1.9335z^{-1} + 0.9345z^{-2}}$	17.3
190 (22-24)	$H(z) = \frac{1 - 1.5047z^{-1} + 0.6428z^{-2}}{1 - 2.0833z^{-1} + 1.2164z^{-2} - 0.1302z^{-3}}$	22.3
191 (24-26)	$H(z) = \frac{1 - 1.5703z^{-1} + 0.6566z^{-2}}{1 - 1.9149z^{-1} + 0.9168z^{-2}}$	16.6

In order to concentrate on the effects of S/A, all errors other than S/A and receiver and satellite clock noise were set to zero. In addition, the user position was constant and a fixed satellite constellation was chosen (figure 10). The horizontal dilution of precision (HDOP) is 2.92, the east dilution of precision is 1.17, and the north dilution of precision is 2.68. Four thousand data points were taken to show the distribution of the horizontal position errors.

The first scenario was designed to verify the simulation model. Pseudorange errors were taken from a normal distribution with zero mean and a standard deviation of 5 meters. As would be expected, the scatter plot of position errors shows them to be approximately Gaussianly distributed (figure 11). It is also comforting to see that the error distribution lies along the line of highest dilution of precision.

With the model thus verified, the two filters derived from the S/A waveforms were used to perturb the pseudoranges. The results using the ARMA(2,2) are given in figure 12 and the results from the ARMA(3,2) are given in figure 13. The position errors again appear to be Gaussianly distributed. This is not surprising. By placing S/A on the four pseudoranges and then computing position, one is essentially convolving four random variables. The central limit theorem states that as N increases, the convolution of N random variables will approach a Gaussian distribution. For short periods of time, on the order of one hour, the effect of S/A on the position error is quite different from the effects of the Gaussian noise, however. The S/A position tends to "wander" around the true user position.

The two dimensional root-mean-squared (2D-rms) position errors for figures 12 and 13 are 93.4 and 157.5 meters, respectively. The average HDOP for a 21-satellite constellation is approximately 1.8 (for an elevation mask angle of 5 degrees). Scaling the simulation results with respect to the average HDOP results in 2D-rms position errors of 57.6 and 97.1 meters, respectively.

SUMMARY AND CONCLUSIONS

A signal model is presented which possesses the capability to generate typical errors in pseudorange measurements. The effects of ionospheric delay, tropospheric delay, and URE are detailed along with the models used to generate them. Two S/A models are derived from actual GPS data using System Identification Theory. The resulting signal model is then used to investigate the effects of S/A on position errors. The results satisfy both theory and intuition.

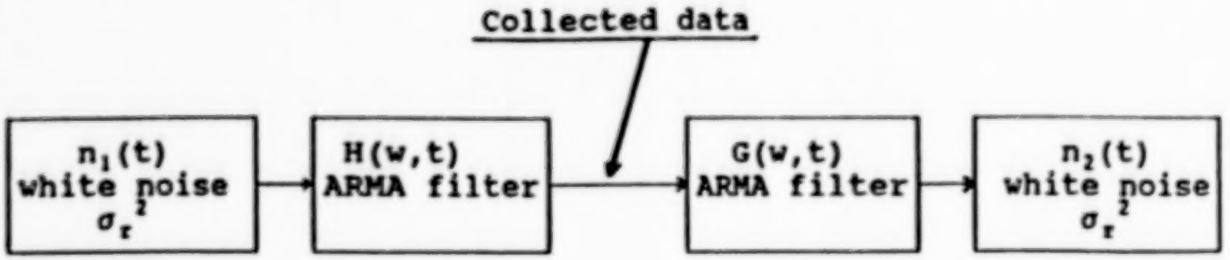
The short term effect of S/A on the horizontal position error is significantly different from the effect of Gaussian noise. Therefore, it is recommended that the System Identification technique outlined in this paper be implemented in GPS receiver test equipment. One principle advantage, as discussed previously is that the test engineer is not limited to a few waveforms of collected S/A. Given the properly identified model, virtually an infinite amount of statistically equivalent S/A data may be generated to exercise the GPS receiver navigation algorithms.

REFERENCES

1. Kelly, R. J., van Graas, F., and Kuhl, M. R.: Improved Effectiveness of GPS RAIM through Ridge Regression Signal Processing. Proceedings of the 2nd International Satellite Division Meeting of the ION, Colorado Springs, CO, September 27-29, 1989.

2. Kalafus, R. M., Van Dierendonck, A. J., Pealer, N. A.: Special Committee 104 Recommendations for Differential GPS Service. Navigation: Journal of the Institute of Navigation, Vol. 33, No. 1, Spring, 1986.

3. Braasch, M. S.: A Signal Model for the NAVSTAR Global Positioning System. Proceedings of the 2nd International Satellite Division Meeting of the ION, Student Session, Colorado Springs, CO, September 27-29, 1989.



$$H(w,t) = G^{-1}(w,t)$$

$$\sigma_r^2 = \text{variance}$$

Figure 1. Concept behind System Identification Theory.

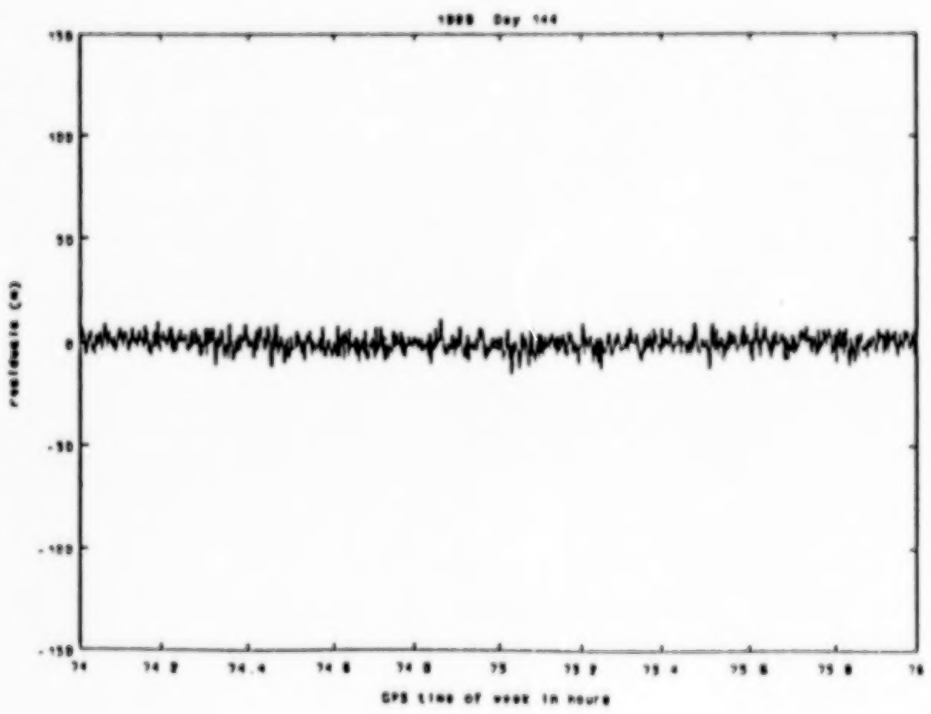


Figure 2. Pseudorange residuals (S/A turned off).

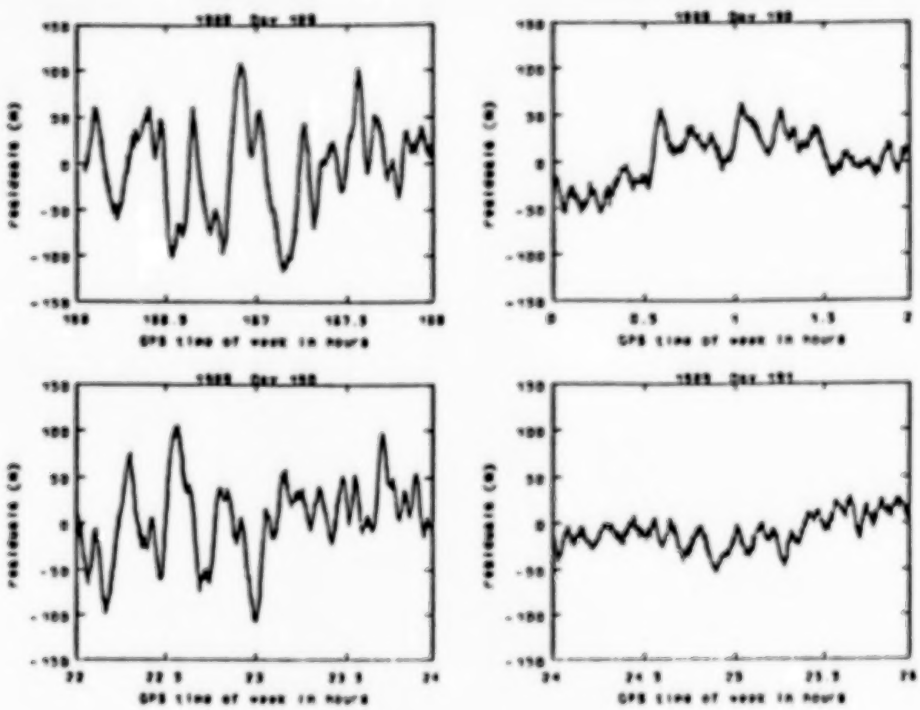


Figure 3. Pseudorange residuals.

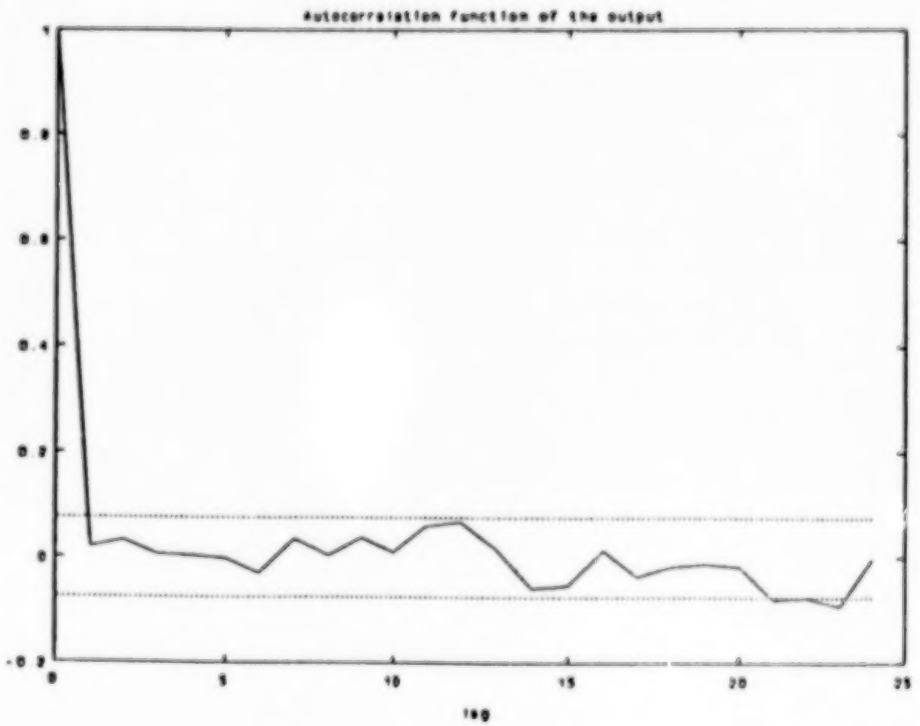


Figure 4. Autocorrelation function of the model output.

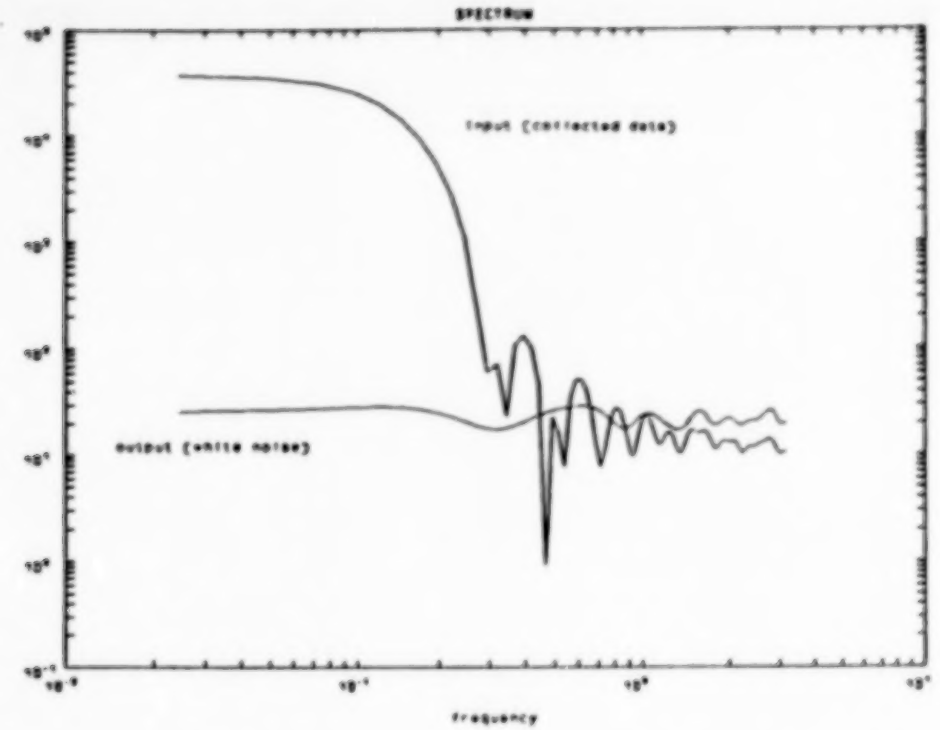


Figure 5. Power spectra of model input and output.

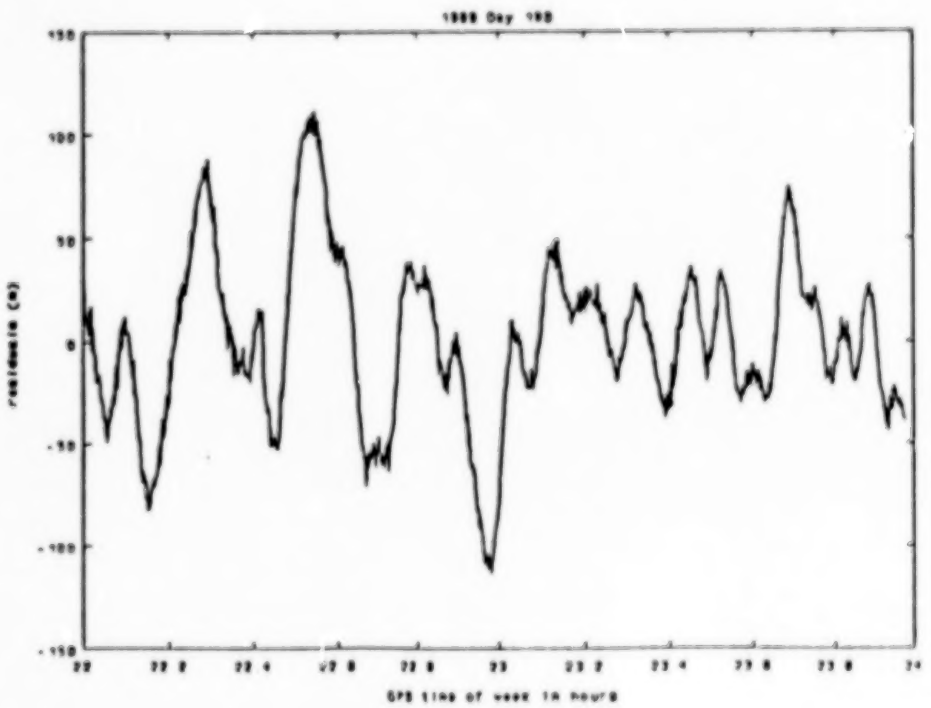


Figure 6. Collected S/A data.

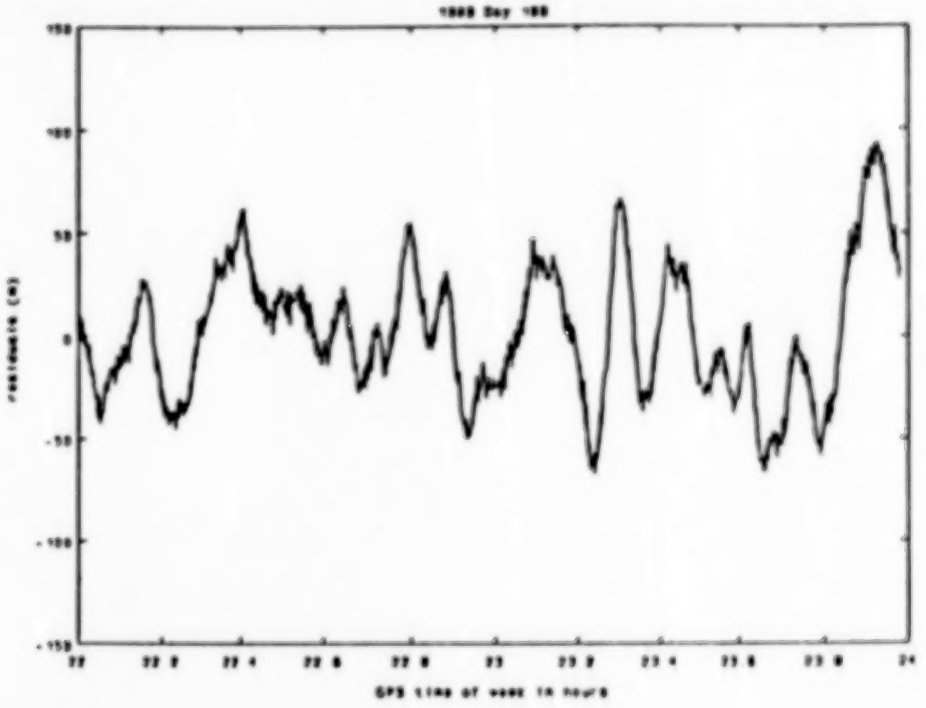


Figure 7. Generated S/A data.

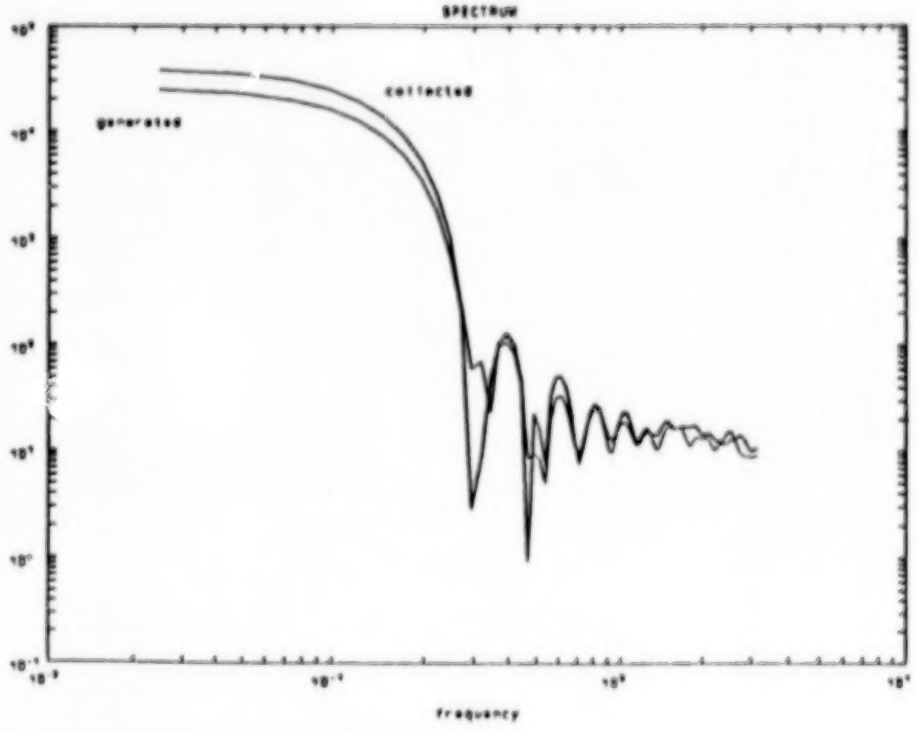


Figure 8. Power spectra of collected and generated data.

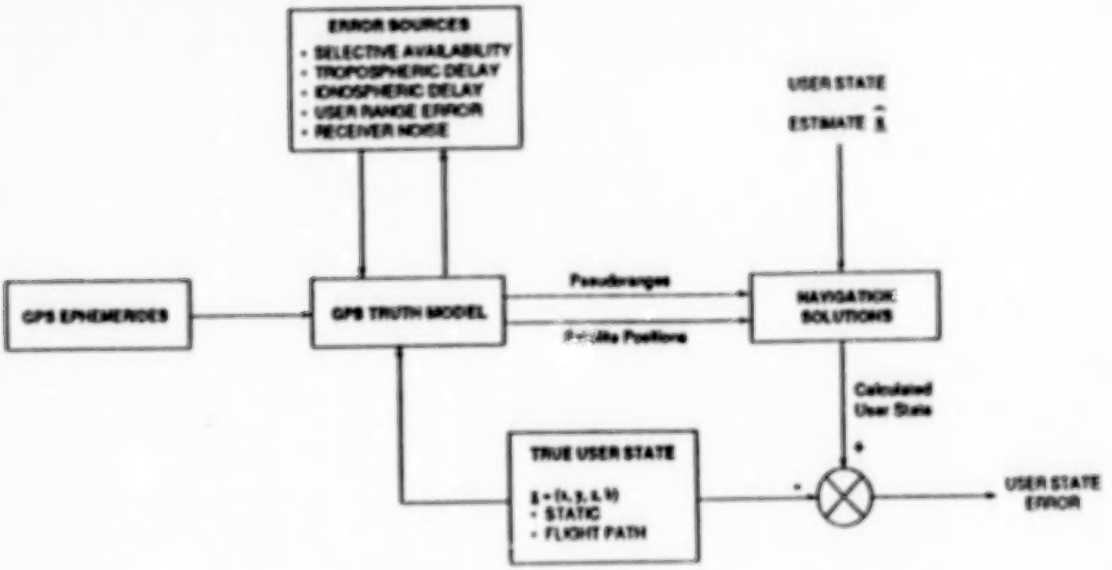


Figure 9. GPS signal model block diagram.

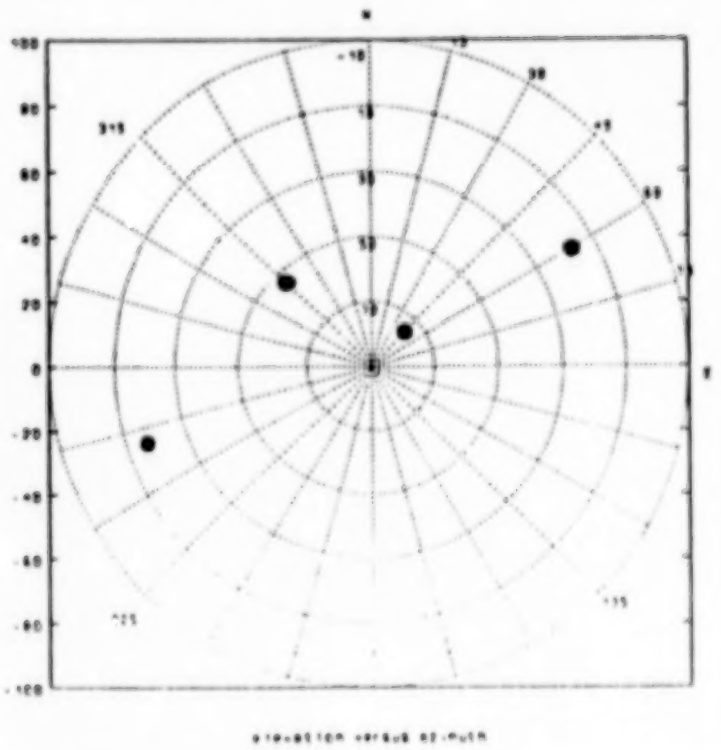


Figure 10. Satellite geometry

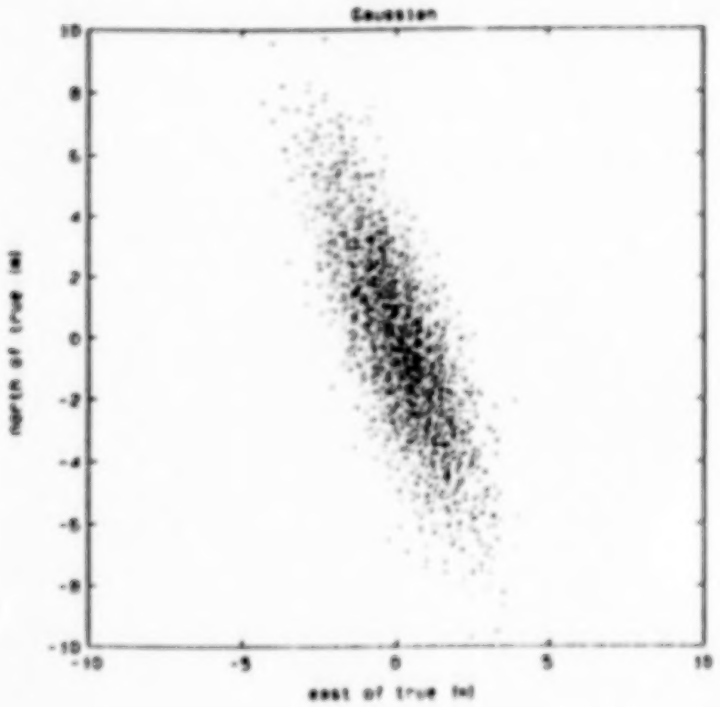


Figure 11. Position errors with Gaussian noise on pseudoranges.

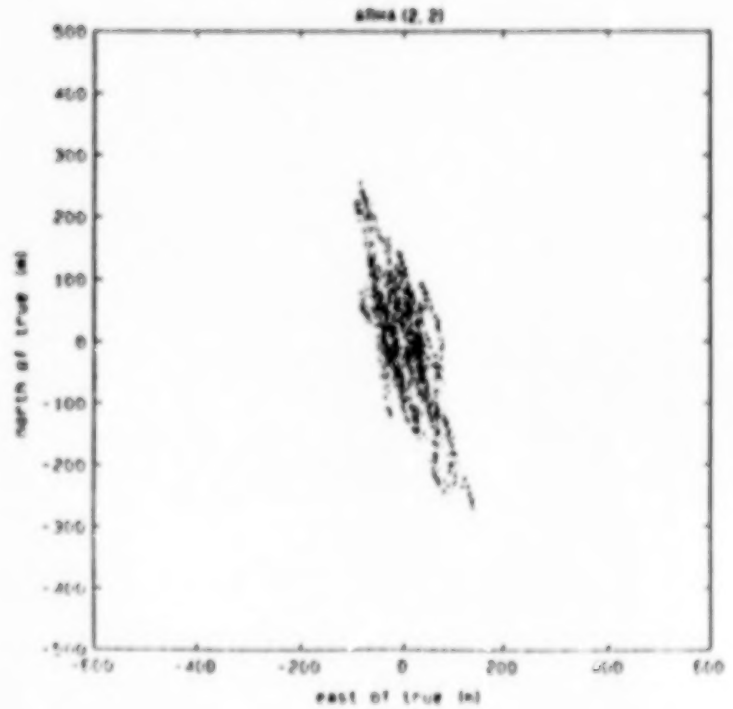


Figure 12. Position errors with ARMA(2,2) generated S/A on pseudoranges.

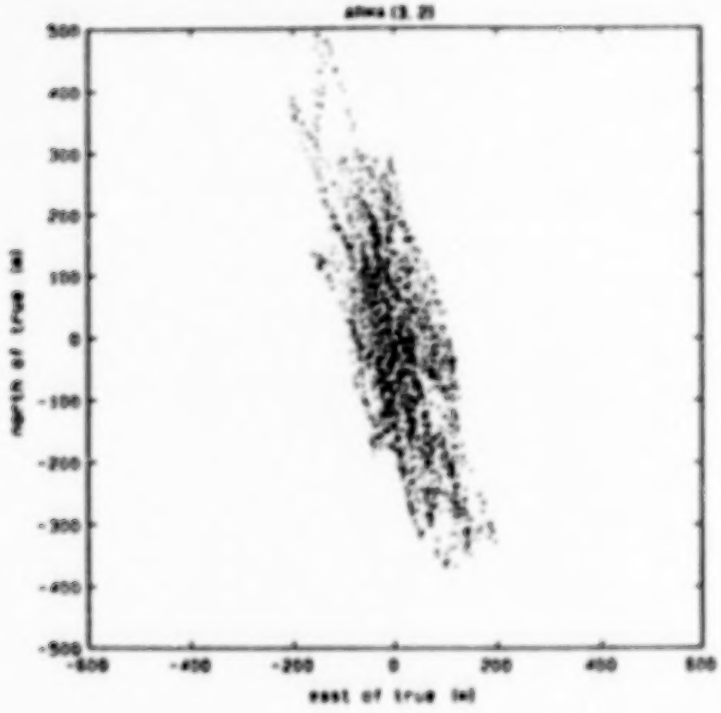


Figure 13. Position errors with ARMA(3,2) generated S/A on pseudoranges.

ANALYZING THOUGHT-RELATED ELECTROENCEPHALOGRAPHIC DATA USING NONLINEAR TECHNIQUES

Trent Skidmore
Ohio University
Athens, Ohio

SUMMARY

A unique method is presented for collecting, studying and interpreting thought-related electroencephalogram (EEG) data. The use of a chaos-based nonlinear analysis technique is shown to be promising in providing insight into relating conscious thought to specific EEG data. A discussion of the practical limitations of this technique is also included.

INTRODUCTION

The current primary use of the EEG is in clinical applications. However, because the EEG involves conceptually simple instrumentation, is relatively safe, and inherently noninvasive, it is well worth investigating for possible use in other applications. Specifically, if it were possible for a human operator to reliably control some aspect of his/her EEG, a direct human-computer interface could be developed. Thus, the goal of this research is to determine if an operator can, with their eyes open, consciously control some EEG feature so as to allow a computer to detect one or more bits of information.

In order to achieve this goal, the laboratory set-up shown in figure 1 was created. The host computer is a Macintosh II utilizing National Instruments' LabVIEW software. The interface board is a National Instruments NB-MIO multifunction A/D converter. The EEG amplifier is a Grass Model 8-16C EEG Instrument. The electrodes are 9 mm silver-silver chloride contacts with collodion adhesive and EEG gel.

The EEG data to be analyzed were obtained during what is called a "Think Math" test. During this test, the subject was instructed to think of a simple math problem (such as 9 x 8 = 72) whenever the light-emitting diode (LED) was turned on. The periods when the light is flashing are called Thought-times. During the off-times of the LED (Gap-times), the subject was instructed not to think about mathematics or multiplication. This on/off "Think Math" process was repeated for several minutes with the LED flashing periodically two out of every eight seconds. The data analyzed in this paper were recorded over what is referred to as the "F3 region with respect to Cz" in the 10-20 International System of Electrode Placement (figure 2). Data were sampled at 128 samples per second (T = 7.8125 ms), well above the Nyquist minimum for data whose highest significant frequency component appears around 15 Hz.

CHAOS ANALYSIS OF EEG DATA

The application of chaos theory to biological systems is not a new concept, nor is the study of the chaotic behavior of the EEG. However, the idea of applying chaos theory to determine if the EEG can be used as a thought/non-thought discriminator appears to be new. A fundamental technique in chaos theory is the reinterpretation of nonlinear time signals as multidimensional phase-space geometric objects. The goal is to determine if there exists a "strange attractor" in the phase-space diagram of the multidimensional data. A two-dimensional phase-space diagram is generated by plotting the original data sequence, $x(nT)$, against a time-shifted version of the same data, $x(nT + \tau)$. Reference 1 suggests that a time shift (τ) be used which is equal to one-quarter of the period of the most significant frequency component in the EEG signal. The reasoning for this value is based on the "mutual information content" of a signal. For the analysis in this paper, a time shift of 23.475 milliseconds was chosen. This corresponds to a dominant frequency of 10.6667 Hz. This is intuitively pleasing since the 10.6667 Hz component is in the middle of the alpha region, considered by some EEG researchers as the most important EEG frequency region.

QUANTIFYING THE STRANGE ATTRACTOR

Figure 3 shows a phase-space diagram for $x(nT + \tau)$ versus $x(nT)$ for 4 seconds of a 32 second "Think Math" test. In order to quantify the chaotic strange attractor generated during this test, the Standard Deviation/Moment of Inertia (SDMI) transformation was developed. This transformation consists of first finding the point-by-point moment of inertia about the origin of the phase-space "cloud." This is denoted by

$$I_x(nT) = x^2(nT) + x^2(nT + \tau).$$

The purpose of this transformation is to convert the two-dimensional phase-space data into a positive scalar whose weight is a function of the original signal and the time-shifted version. After this transformation, a moving-average, standard deviation calculation is performed. This is computed as follows:

$$\sigma(k) = (1/m) \sum_{n=1+mk}^{(k+1)m} (I_x(nT) - \mu_k)^2 \quad \text{where:}$$

$$\mu_k = \sum_{n=1+mk}^{(k+1)m} (1/m) I_x(nT); \quad k=0,1,2,...,(N/m)-1$$

where N is the total length of the data sequence and m is the standard deviation window length (64 points).

This calculation computes a windowed standard deviation of the moment of inertia of the phase-space "cloud" twice per second (every 64 points). Figure 4 shows the results of this

analysis by plotting $\sigma(k)$ and the thought-times versus time. Note that this entire run contains four two-second thought times and four six-second gap (non-thought) times.

CONCLUSIONS

For this subject, it is clear that the SDMI analysis produces a significant peak in the standard deviation plot during the times of transition between thinking and not thinking. In attempting to generalize the results shown here, the entire test was repeated on several other subjects. These subsequent tests, which utilized either visual thought cueing or audio cueing, yielded results very similar to that presented here. Further investigation of the raw EEG did indicate the possibility that unwanted artifact (either eye or muscle movement) contributed to the peaks generated by the SDMI analysis. Thus, some caution is given in interpreting the SDMI peaks as purely thought-related activity. While this research is still in its infancy, it is apparent from these results that this technique may be useful for understanding thought-related EEG data.

REFERENCES

1. Mayer-Kress, G., and Layne, S.: Dimensionality of the Human Electroencephalogram. Proceedings of the New York Academy of Sciences Conference on Perspectives in Biological Dynamics and Theoretical Medicine, Bethesda, MD, April 9-11, 1986.

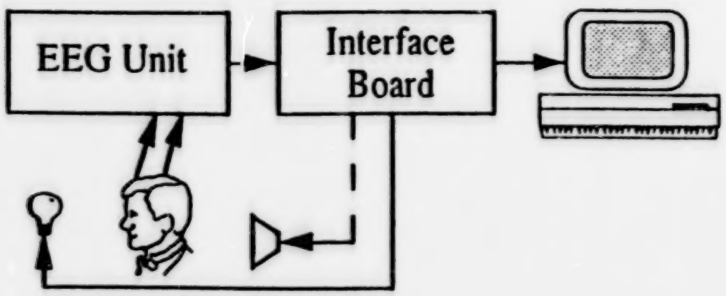


Figure 1. EEG Workstation using a Macintosh II and a Grass Model 8-16C EEG Unit.

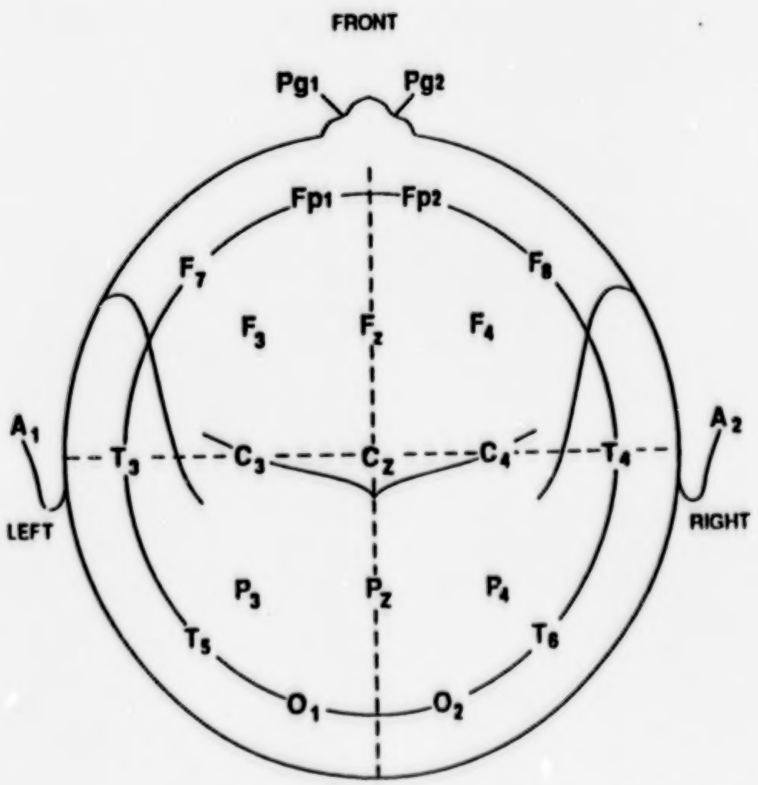


Figure 2. The 10-20 International System of Electrode Placement.

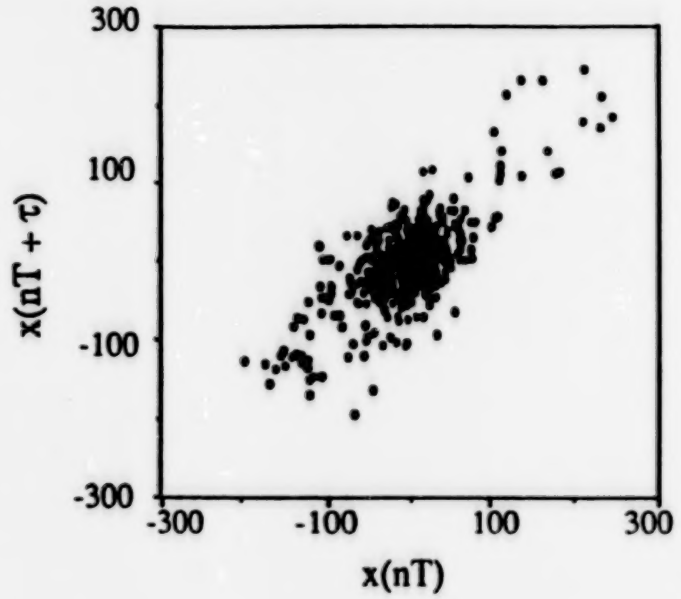


Figure 3. Chaotic Phase-Space "Cloud" showing x(nT + τ) versus x(nT) for four seconds of data.

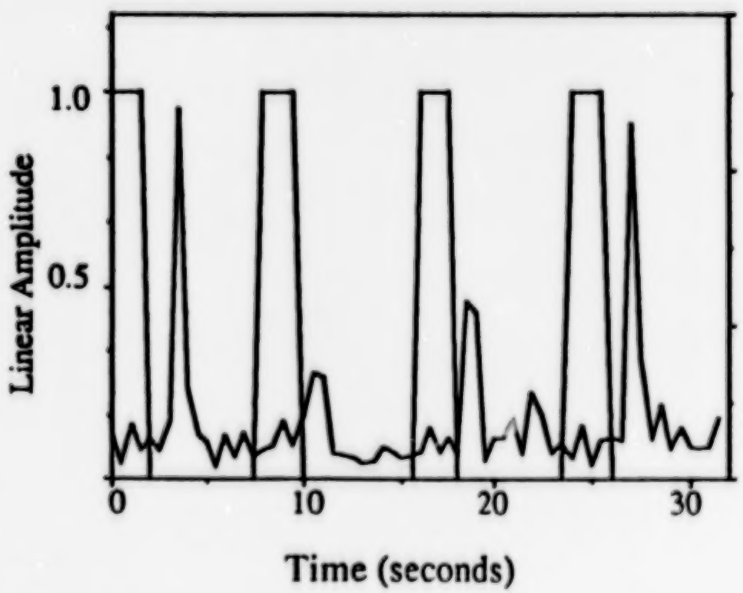


Figure 4. Thought-Times and the SDMI EEG data for a 32 second "Think Math" run.

INTEGRATED INERTIAL/GPS

Paul Kline and Frank van Graas
Ohio University
Athens, Ohio

SUMMARY

The presence of failures in navigation sensors can cause the determination of an erroneous aircraft state estimate, which includes position, attitude, and their derivatives. Aircraft flight control systems (AFCS) rely on sensor inputs to determine the aircraft state. In the case of integrated Inertial / NAVSTAR Global Positioning System (GPS), sensor failures could occur in the on-board inertial sensors or in the GPS measurements. The synergistic use of both GPS and INS allows for highly reliable fault detection and isolation of sensor failures. Integrated Inertial/GPS is a promising technology for the High Speed Civil Transport (HSCT) and the return and landing of a manned space vehicle.

BACKGROUND

The NAVSTAR Global Positioning System is expected to be operational by the year 1993. Presently, ten experimental satellites (Block I) and seven operational satellites (Block II) have been successfully launched and are undergoing tests. Navigation with the GPS is based on pseudorange and integrated Doppler shift measurements with respect to GPS satellites. In combination with ephemeris information, transmitted by the satellites, a receiver can determine its position and velocity with respect to the satellites with a positional accuracy of 16 meters (Precise Positioning Service) and with a velocity accuracy of better than 0.1 meter per second. Differential GPS can result in navigational accuracies of better than 3 meters for pseudoranging and well within 0.2 meter using integrated carrier phase measurements. The GPS satellite signals can also be used to determine aircraft attitude by making interferometric measurements with a minimum of three separate antennas (ref. 1). GPS interferometric attitude determination is expected to be within 0.1 degree for pitch, roll, and yaw angles.

Because of the capabilities described above, the GPS can play an important role in the reliable state estimation of aircraft. Specifically, an integrated Inertial/GPS system could provide continuous, high-accuracy attitude information, three-dimensional position and velocity, and time. The system could also provide a high level of fault tolerance, since GPS and inertial measurements provide similar information.

The overall objective of an integrated Inertial/GPS system is to provide reliable aircraft state estimates in the presence of possible failures in the GPS measurements and the on-board inertial sensors. This involves the design of fault tolerant algorithms, implementation and flight demonstration of the integrated Inertial/GPS system. Previous research at Langley Research Center demonstrated the capability to provide dual fail-operational performance for a skewed array of inertial sensors (ref. 2). The redundant strapdown inertial measurement unit

(RSDIMU) used for this project will also be used for the laboratory and flight demonstration of Inertial/GPS techniques.

OVERVIEW OF RESEARCH

Current efforts are focused on the determination of the effective GPS data rate which involves GPS receiver tracking loop design and simulation, and on fault detection and isolation of GPS measurement errors. In the area of receiver autonomous fault detection, a novel signal processing technique has been applied to improve the effectiveness of fault detection algorithms in the presence of near collinearity (ref. 3).

The GPS receiver simulation program as presented in reference 1 is being extended to address the effect of tracking loop bandwidths of the code and carrier tracking loops on the effective GPS data rate. Figure 1 shows the block diagram of an integrated Inertial/GPS in combination with the AFCS. Both GPS and inertial data are combined by the estimator, which in turn provides the aircraft state estimates to the flight control system. The GPS data rate is required to facilitate the integration of GPS data into the aircraft flight control system, and to determine fault detection capabilities based on GPS attitude information.

Also continuing is the implementation of the integrated Inertial/GPS test bed. The RSDIMU system is currently available for data collection in the laboratory environment. The inertial system will be combined with a 4-channel GPS receiver. At the same time, provisions are being made to install the integrated system in Ohio University's Douglas DC-3 research aircraft, N7AP.

REDUNDANT STRAPDOWN INERTIAL MEASUREMENT UNIT

The inertial sensor used for this project is a Redundant Strapdown Inertial Measurement Unit (RSDIMU). A block diagram of the RSDIMU is shown in figure 2 (based on ref. 4). The inertial sensors consist of four two-degree-of-freedom accelerometers and gyroscopes arranged in two semi-octahedral arrays. These redundant arrays are interfaced to electronics boxes which contain analog-to-digital converters. The sensor data are processed by two flight computers. The magnetic tape drive is used for in-flight program loading. Also used during system testing is a printer for real time data output and a control panel to operate the system and to inject failures into the measurement data. Power to the inertial sensors is provided through a power back-up unit, which allows the sensors to operate for about 20 minutes in the absence of aircraft power.

The performance of the inertial sensors would support navigation accuracies on the order of 2 nautical miles per hour. However, the inertial system is only required to provide short-term navigation data. The GPS will provide the long-term navigation stability for this system.

RSDIMU NAVIGATION MODE PROGRAM FLOW

The RSDIMU has several modes of operation, including coarse and fine alignment, and navigation. Figure 3 shows the flow diagram for the navigation mode. Each navigation cycle

is initiated by a sensor interrupt. The raw sensor data are then scaled and converted to floating point representation. Next, fault detection and isolation routines are carried out, and some data are fed back to the IMU sensors for corrections. Finally, attitude information is output to the flight control system and the navigation solution is updated. The subsequent sensor interrupt initiates the next run through the navigation update cycle. The system operates at an update rate of 64 updates per second.

GPS RECEIVER

A four-channel GPS receiver employing continuous carrier tracking is used for this project. Figure 4 shows a simplified block diagram of the GPS receiver tracking loops. The bandwidth of the code tracking loop is as small as 0.01 Hertz when the receiver is in continuous tracking. The carrier tracking loop bandwidth is 32 Hertz and is used to update the code tracking loop every millisecond to correct for dynamics which cannot be tracked by the code loop. The GPS carrier wavelength at 1575.42 MHz is approximately 19 cm, which provides a tracking resolution of better than 1 cm. In addition, the carrier tracking loop is a second order system which has less than 20 degrees of phase lag at line-of-sight accelerations of up to 40 m/s^2, which corresponds to a tracking error of less than 1 cm. The continuous carrier tracking is essentially a dead-reckoning input, similar to Doppler radar.

APPLICATIONS

The integration of GPS and INS has been targeted for several possible applications. These applications actually define the overall system requirements. One application is the High Speed Civil Transport (HSCT). GPS/INS can be used for heading and attitude determination. The return of a Space Shuttle from orbit or a Manned Return Vehicle (MRV) from a space station is another application. The inertial measurement would provide the short-term stability required by the flight control system. Differential GPS would be used in the terminal area to greatly improve landing accuracy and to enable the use of candidate landing sites throughout the world without the need for a dedicated landing system installation.

CONCLUSIONS

The synergistic integration of inertial measurements and the GPS operated stand-alone or in the differential mode has many promising applications in the areas of reliable aircraft state estimation and for use in terminal and landing operations. The RSDIMU test bed is operational in the laboratory environment, and is being integrated with a GPS receiver to implement and evaluate the above applications and to develop fault detection and isolation algorithms.

REFERENCES

1. McIntyre, D. S.: GPS Effective Data Rate Characterization with Application to Integrated GPS/INS Attitude and Heading Determination. Master's Thesis, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, November 1989.

2. Morrell, F. R., Bailey, M. L., and Motyka, P. R.: Flight Demonstration of Redundancy Management Algorithms for a Skewed Array of Inertial Sensors. Proceedings of the AIAA/AHS/ASEE Aircraft Design and Operations Meeting, AIAA Paper 88-4434, Atlanta, GA, September 7-9, 1988.

3. Kelly, R. J., van Graas, F., and Kuhl, M. K.: Improved Effectiveness of GPS RAIM through Ridge Regression Signal Processing. Proceedings of the 2nd International Satellite Division Meeting of the ION, Colorado Springs, CO, September 27-29, 1989.

4. Bryant, W. H. and Morrell, F. R.: Description of a Fail-Operational Redundant Strapdown Inertial Measurement Unit for Avionics Research. Proceedings of the 4th ROLM Computer Users Group Meeting, San Diego, CA, February 1981.

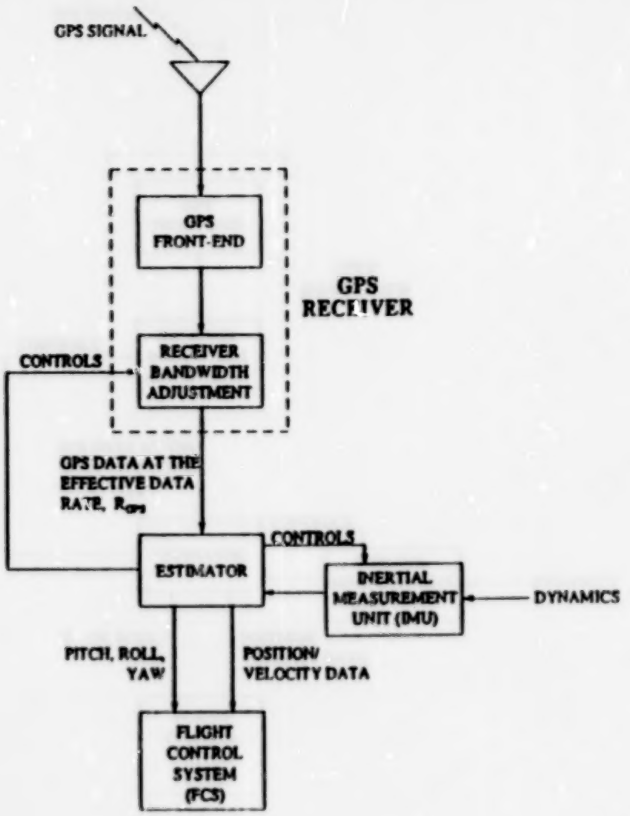


Figure 1. Integrated Inertial/GPS in combination with the Aircraft Flight Control System.

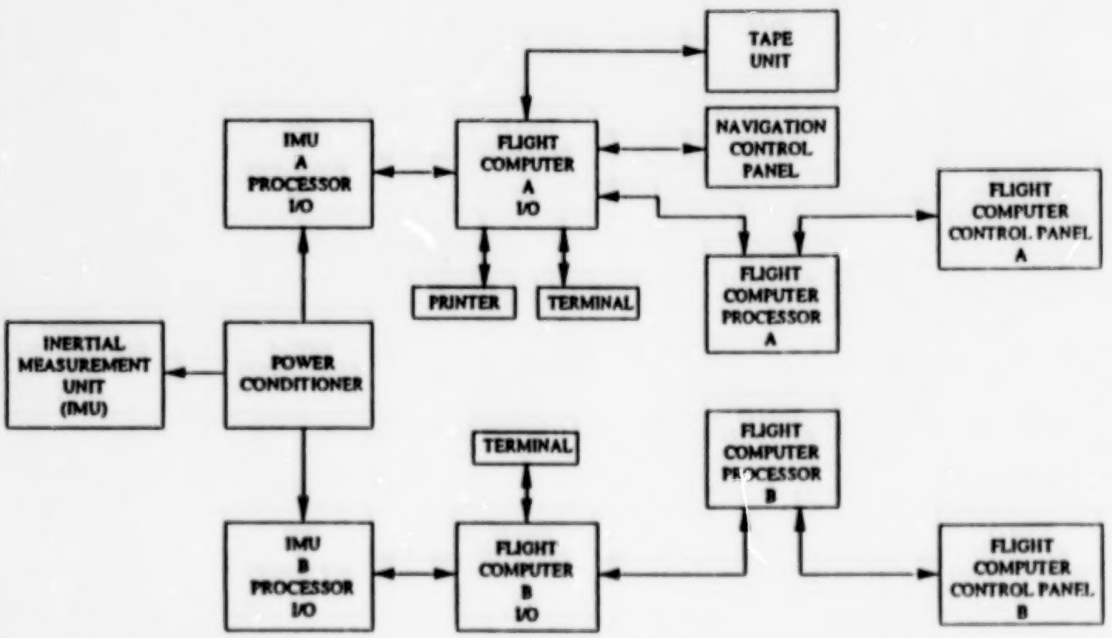


Figure 2. Block diagram of the Redundant Strapdown Inertial Measurement Unit (RSDIMU) (based on ref. 4).

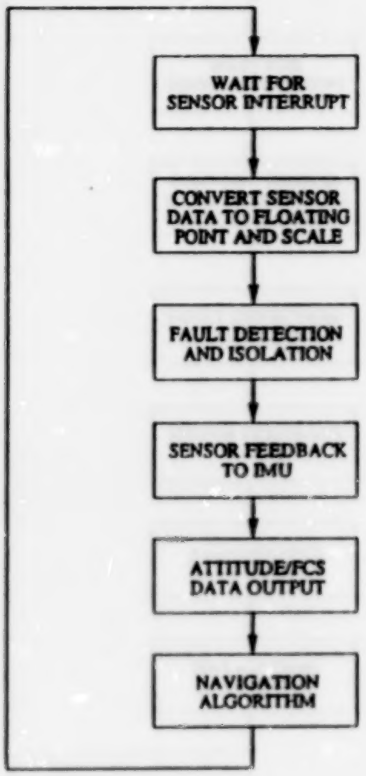

WAIT FOR SENSOR INTERRUPT
CONVERT SENSOR DATA TO FLOATING POINT AND SCALE
FAULT DETECTION AND ISOLATION
SENSOR FEEDBACK TO IMU
ATTITUDE/FCS DATA OUTPUT
NAVIGATION ALGORITHM

Figure 3. RSDIMU navigation mode flow diagram.

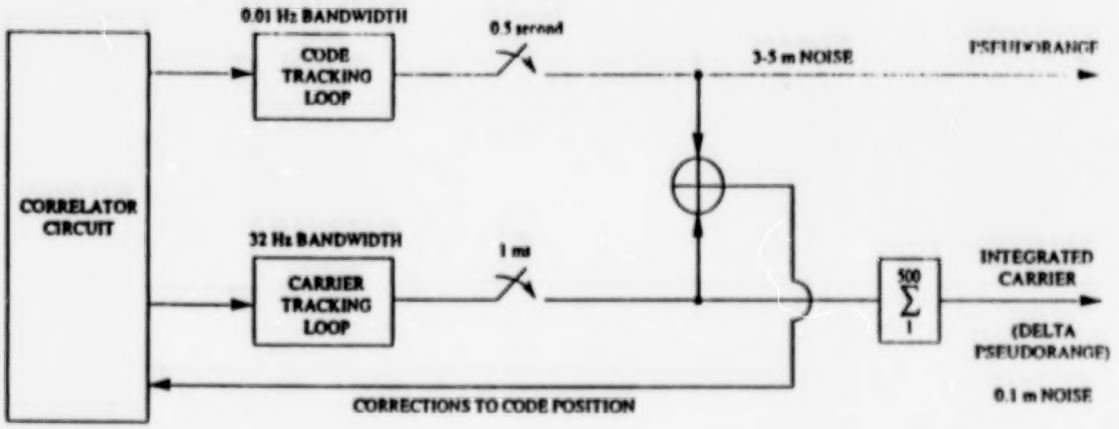


Figure 4. GPS receiver simplified block diagram of the code and carrier tracking loops.

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT PRINCETON UNIVERSITY, 1989-1990

Professor Robert F. Stengel
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey

SUMMARY OF RESEARCH

The Air Transportation Technology Program at Princeton University, a program emphasizing graduate and undergraduate student research, proceeded along six avenues during the past year:

- Microburst Hazards to Aircraft
- Machine-Intelligent, Fault-Tolerant Flight Control
- Computer-Aided Heuristics for Piloted Flight
- Stochastic Robustness of Flight Control Systems
- Neural Networks for Flight Control
- Computer-Aided Control System Design

This research has resulted in a number of publications, including theses, archival papers, and conference papers. An annotated bibliography of publications that appeared between January, 1989, and June, 1990, appears at the end of this report. The research that these papers describe was supported in whole or in part by the Joint University Program, including work that was completed prior to the reporting period.

Severe downdrafts and resulting high velocity outflows caused by microbursts present a significant hazard to aircraft on takeoff and final approach. *Microbursts*, which are often associated with thunderstorm activity, also can occur in the vicinity of dissipating convective clouds that produce no rainfall at ground level. Microburst encounter is a rare but extremely dangerous phenomenon that accounts for one or two air carrier accidents and numerous general aviation accidents each year (on average). Conditions are such that an aircraft's performance envelope may be inadequate for safe penetration unless optimal control strategies are known and applied.

Our current wind shear research focuses on avoiding wind shear during transport aircraft operations, as well as on developing a cockpit simulation of wind shear encounter. Graduate student Alex Stratton is developing an expert system for wind shear avoidance that extends the FAA Microburst Windshear Guidelines to account for temporal and spatial variations in the evidence that wind shear is present [1]. The approach being taken is to develop a Bayesian Belief Network that relates information gathered from many sources to determine the probability of encountering a microburst on the intended flight path. Our principal objectives are to develop methods for assessing the likelihood of wind shear encounter (based on real-time information in the cockpit), for deciding what flight path to pursue (e.g., abort, go-around, normal climbout, or glide slope), and for using the aircraft's full potential to combat wind shear. This study requires the definition of deterministic and statistical techniques for fusing internal and external information, for making "go/no-go" decisions, and for generating commands to the aircraft's autopilot and flight directors in automatic and manually controlled flight.

Graduate student Sandeep Mulgund has begun the development of a fixed-base cockpit simulator for microburst studies. The simulation incorporates a cockpit station with manual input devices and a graphical display of instruments and the "out-the-window" view. Currently the simulator is programmed to simulate a twin-jet transport aircraft and a single-engine general aviation airplane. Displays for the "out-the-window" view and the control panel devices are generated entirely by computers and are presented on color cathode-ray tubes. The external view is generated and displayed by a Silicon Graphics IRIS 3020 Workstation. A central computing unit performs dynamic and control calculations, accepts analog inputs, drives the panel displays, and commands the external view. It is a special-purpose Multibus computer employing parallel 80286 processors and special-purpose graphics boards, which are controlled by an IBM PC-AT computer.

Undetected system failures and/or inadequately defined recovery procedures have contributed to numerous air carrier incidents and accidents. The infamous DC-10 accident at Chicago's O'Hare Airport, in which loss of an engine pod, subsequent loss of subsystems, and asymmetric wing stall led to disaster, provides a prototype for the kind of tragedy that could be averted by intelligent flight control systems. (An intelligent control system is one that uses artificial intelligence concepts, e.g., an expert systems program, to improve performance and failure tolerance.) Although many methods of modern control theory are applicable, the scope of the problem is such that none of the existing theories provides a

complete and practical solution to the problem. At the same time, heuristic logic may be applicable, but it has yet to be stated in satisfactory format.

Graduate student David Handelman developed a knowledge-based reconfigurable flight control system that could be considered a prototype for a failure-tolerant control system constructed with available hardware. The knowledge-based flight control system is specified initially and tested using the LISP programming language. When desired logic is determined, the corresponding Pascal code is generated automatically for real-time implementation (which has been demonstrated in a laboratory environment) [2,3]. Chien Huang addressed the failure-analysis problem in his Ph.D. thesis [4], and he developed a method for online control reconfiguration based on proportional-integral implicit model-following [5]. Parvadha Suntharalingam examined three alternatives for model-following control in her M.S.E. thesis [6].

Helping a pilot make quick decisions under high workload conditions is important for aircraft missions of all types. In research principally supported by an Army/Navy grant but reported at numerous quarterly reviews of the Joint University Program, Brenda Belkin has developed an expert system of expert systems called AUTOCREW. In her M.S.E. thesis [7], Ms. Belkin uses the paradigm of a hypothetical aircraft crew to facilitate the assignment of tasks, rules, and data within parallel knowledge bases. AUTOCREW performs a cyclical search in which the Director expert system, the electronic analog of the aircraft commander, establishes goals that invoke the crew-member expert systems. The crew members then perform such tasks as observation, monitoring, and control in response to continuing needs as well as special requests from the Director. Particular attention was paid to the navigation expert system, which manages diverse sensor inputs under varied trajectory and failure scenarios [8].

Control system robustness is defined as the ability to maintain satisfactory stability or performance characteristics in the presence of all conceivable system parameter variations. While assured robustness may be viewed as an alternative to gain adaptation or scheduling to accommodate known parameter variations, more often it is seen as protection against uncertainties in plant specification. Consequently, a statistical description of control system robustness is consistent with what may be known about the structure and parameters of the plant's dynamic model. Guaranteeing robustness has long been a design objective of control system analysis, although in most instances, insensitivity to parameter variations has been treated as a deterministic problem. Graduate students Laura Ryan Ray and Chris Marrison are investigating a simple numerical procedure for estimat-

ing the *stochastic robustness* of control systems. Monte Carlo evaluation of the system's eigenvalues allows the *probability of instability* and the related *stochastic root locus* to be estimated. This analysis approach treats not only Gaussian parameter uncertainties but non-Gaussian cases, including uncertain-but-bounded variations [9-13,14,15]. Trivial extensions of the procedure admit alternate discriminants to be considered. Thus, the probabilities that stipulated degrees of instability will be exceeded or that closed-loop roots will leave desirable regions also can be estimated. Results are particularly amenable to graphical presentation.

There is growing interest in the use of neural networks for computational decision-making and control, brought about by the advent of small, fast, inexpensive computers. The neural network paradigm offers a potentially attractive structure for flight control systems that adapt to changing flight conditions and system failures, but much is to be learned about the practicality of such an approach. Graduate student Dennis Linse has begun to examine this potential. Current research focuses on the use of two types of neural networks (feed-forward back-propagation network and cerebellar model articulation controller) in an adaptive nonlinear controller based upon feedback linearization [16,17]. In this application, a suite of six neural networks represents the nonlinear aerodynamic/thrust model of the aircraft, and the model can be trained both before and during operational flight.

Graduate student Subrata Sircar has begun to examine concepts for the next generation of computer-aided control system design. The principal goal is to add knowledge-based features to the design logic, such as background autonomous anticipatory operation, which will keep the computer busy performing useful work while the user is evaluating the next course of action, and new graphical output paradigms. The control designer should be able to enter the control equations directly, with as few restrictions and special conventions as possible. The approach will use the Control Equation Parser (CEP), previously developed at Princeton, which accepts symbolic expressions and performs the needed translations using the LISP programming language. The present embodiment of CEP has three parts (preprocessor, parser, and postprocessor), and it runs on both the Symbolics 3670 LISP Machine and the NeXT Computer.

ANNOTATED REFERENCES OF 1989-90 PUBLICATIONS

1. R. Stengel and D.A. Stratton, An Expert System for Wind Shear Avoidance, *Engineering Applications of Artificial Intelligence*, Vol. 2, No. 3, Sept 1989, pp. 190-197. (also *Proceedings of the 1989 American Control Conference*, Pittsburgh, June 1989, pp. 349-354.)*

A study of intelligent guidance and control concepts for protecting against the adverse effects of wind shear during aircraft takeoffs and landings is being conducted, with current emphasis on developing an expert system for wind shear avoidance. Principal objectives are to develop methods for assessing the likelihood of wind shear encounter (based on real-time information in the cockpit), for deciding what flight path to pursue (e.g., takeoff abort, landing go-around, or normal climbout or glide slope), and for using the aircraft's full potential for combating wind shear. This study requires the definition of both deterministic and statistical techniques for fusing internal and external information, for making "go/no-go"decisions, and for generating commands to the aircraft's autopilot and flight directors for both automatic and manually controlled flight. The program has begun with the development of the Wind Shear Safety Advisor, an expert system for pilot aiding that is based on the FAA Wind Shear Training Aid, a two-volume manual that presents an overview, pilot guide, training program, and substantiating data that provides guidelines for this initial development. The Wind Shear Safety Advisor expert system currently contains over 200 rules and is coded in the LISP programming language.

2. D. Handelman and R. Stengel, Combining Expert System and Analytical Redundancy Concepts for Fault-Tolerant Flight Control, *J. Guidance, Control, and Dynamics*, Vol. 12, No. 1, Jan-Feb 1989, pp. 39-45.*

A technique for rule-based fault-tolerant control is presented. The objective is to define methods for designing control systems capable of accommodating a wide range of aircraft failures, including sensor, control, and structural failures. A software architecture that integrates quantitative analytical redundancy techniques and heuristic expert system concepts for the purpose of in-flight, real-time fault tolerance is described. The resultant controller uses a rule-based expert system approach to transform the

Archival Papers*
Conference Papers**
Ph.D. Theses+
M.S.E. Theses++

problem of failure accommodation task scheduling and selection into a problem of search. Control system performance under sensor and control failures using linear discrete-time simulations of a tandem-rotor helicopter's dynamics is demonstrated. It is found that the rule-based control technique enhances existing redundancy management systems, providing smooth integration of symbolic and numeric computation, a search-based decision-making mechanism, straightforward system organization and debugging, an incremental growth capability, and inherent parallelism for computational speed.

3. Handelman, D.A., *A Rule-Based Paradigm for Intelligent Adaptive Flight Control*, Ph. D. Thesis, Princeton University, Report No. 1858-T, April 1989.+

Although fault tolerance always has been an important aspect of aircraft design, the reduced static stability and increased maneuverability of modern aircraft complicate the problem by shortening the amount of time available to detect, identify, and adjust for component failures. This dissertation investigates the use of highly integrated symbolic and numeric processing in real-time knowledge-based systems for enhanced automatic aircraft failure accommodation. A rule-based control technique is proposed whereby procedural activity is attained through the manipulation of declarative expressions. Rules are used to encode common-sense dependencies, to incorporate expert knowledge on specific situations, and to invoke algorithmic mathematical procedures. Task execution occurs as a by-product of search through these knowledge-base rules. Also proposed is a rule-based controller development system that utilizes a high-level symbolic LISP environment for preliminary system design. Automatic LISP-to-Pascal knowledge-base translation is used to provide dramatically increased execution speed and an environment for highly integrated symbolic and numeric computation. The utility of the control technique is demonstrated through the construction of a multi-processor Rule-Based Flight Control System (RBFCS) for a CH-47 tandem-rotor helicopter. The RBFCS is shown to accommodate multiple simulated failures affecting the electrical, hydraulic, and stability augmentation subsystems of the helicopter. It is concluded that declarative rules with embedded procedural code provide a sound basis for the construction of complex control systems.

4. C. Huang, *A Methodology for Knowledge-Based Restructurable Control to Accommodate System Failures*, Ph. D. Thesis, Princeton University, Report No. 1869-T, Aug 1989.†

The problem of dealing with unanticipated failures of dynamic systems is addressed. The solution centers on developing a knowledge-based approach to perform online failure isolation and system reconfiguration. The failure analysis employs inference mechanisms and causal relationships among the devices to generate probable failure models, which are pruned and ranked by the use of heuristic measures. Hypotheses for other probable causes are made if the failure model does not adequately account for the anomalies. After the failure source is identified, the system knowledge base is updated and the impact of failure is assessed. Any degradation of system performance caused by failures is restored by restructuring the control. This issue is studied in depth and addressed by the analysis and design of an implicit-model-following control law. The concepts and methodologies of the approach are illustrated via many examples and further substantiated using a number of failure scenarios involving a tandem-rotor helicopter. The results show that by providing a form of intelligent self-organization, the knowledge-based restructurable control approach potentially is capable of accommodating unanticipated failures.

5. C. Huang and R. Stengel, Restructurable Control Using Proportional-Integral Implicit Model Following, *J. Guidance, Control, and Dynamics*, Vol. 13, No. 2, Mar-Apr 1990, pp. 303-309.*

Studies of a proportional-integral implicit-model-following control law are presented. The research focuses on the ability of the control law to recover the performance of a system with failed actuators or structural damage to its prefailure level. Properties of the implicit-model-following strategy are examined, and conditions for control reconfiguration are stated. The control law is applied to the lateral-directional model of a fighter aircraft, and control restructuring is shown for changes in control and system matrices. It is concluded that the implicit-model-following scheme is a good candidate for control reconfiguration.

6. P. Suntharalingam, *Applications of Computer-Aided Control System Design to Linear Quadratic Model-Following*, M.S.E. Thesis, Princeton University, Report No. MAE-1868-T, Aug 1989.++

This thesis deals with the application of optimal control methods to aircraft flight dynamics. The technique of model-following control has been found to be an effective way of incorporating certain design specifications into the formulation of a linear-quadratic control law. This project focused, in particular, on the behavior of the implicit model-follower and the explicit model-follower. The design criterion for the former controller is the minimization of the difference between plant and model dynamics, and the latter controller is based on minimization of the error between plant and model states. In addition, in an attempt to combine the favorable aspects of both methods, a hybrid model-follower was developed, incorporating features of both implicit and explicit model-followers.

The relative performance of the three controllers was evaluated by examining eigenvalues, time responses, robustness, and reconfiguration possibilities following failure. Results indicate that the explicit model-follower is superior in aspects of robustness and steady-state performance. The implicit model-follower offers specific advantages in the transient response to command input and in the use of lower gains. The hybrid controller combined desirable aspects of the two controllers, yielding the best overall performance.

7. Belkin, B.L., *Cooperative Rule-Based Systems for Aircraft Navigation and Control*, M.S.E. Thesis, Princeton University, Report No. MAE-1856-T, June 1989.++

This thesis proposes a design methodology for the development of multiple cooperating rule-based systems for aircraft. Nine systems, collectively called AUTOCREW, were designed to automate functions and decisions. The organization of tasks is described, details of knowledge-base development and implementation are given, and performance metrics for evaluating the workload of each knowledge base are demonstrated. Several test scenarios were evaluated using an interactive graphical simulation on an IBM PC-AT computer. Software tools developed to aid in high-level design also are described. Results show that these tools facilitate rapid prototyping of a complex system exhibiting knowledge-base cooperation, satisfactory logic flow, and human pilot-AUTOCREW interaction.

Design of one of the component expert systems, the Navigation Sensor Management (NSM) module, was pursued in considerable detail. This problem was chosen because it presented the challenge of designing an expert system from simulation data, that is, from quantitative test results. The NSM Expert was systematically derived from Kalman filter covariance data for simulated missions flown with seven different navigation system types. This development used Analysis of Variance (ANOVA) and the "ID3" algorithm. The function of the NSM Expert was to determine optimal navigation strategies from a set of available sensors based on a root-sum-square metric. Results show that the NSM Expert predicted position accuracy between 65 and 100 percent of the time for a specified "navaid" configuration and aircraft trajectory; hence, this decision-making logic could be incorporated in a scheme for best navaid selection. The systematic nature of the ANOVA/ID3 method makes it broadly applicable to expert system design when experimental or simulation data are available.

8. B. Belkin and R. Stengel, Quantitative Knowledge Acquisition for Expert Systems, presented at the *Space Operations, Applications, and Research Symposium*, Albuquerque, NM, June 1990.**

A common problem in the design of expert systems is the definition of rules from empirical data obtained in system operation or simulation. While it is relatively easy to collect numerical data and to log the comments of human operators engaged in experiments, generalizing such information to a set of rules has not previously been a straightforward task. This paper presents a statistical method for generating the needed rule base from numerical data, motivated by an example based on vehicle navigation with multiple sensors. The specific objective is to design an expert system that selects a satisfactory suite of measurements from a dissimilar, redundant set, given an arbitrary navigation geometry and possible sensor failures.

This paper describes the systematic development of a Navigation Sensor Management (NSM) Expert System from Kalman Filter covariance data. The development method consists of the two statistical techniques: *Analysis of Variance (ANOVA)* and the *ID3 algorithm*. The ANOVA technique indicates whether variation of a problem parameter gives *statistically* different covariance results, and the ID3 algorithm identifies the *relationships* between the problem parameters using probabilistic knowledge extracted from a simulation example set. ANOVA results show that statistically different position accuracies are obtained when different navaids are used, the number of navigation aids is changed, the trajectory is varied, or the performance history is altered. By indicating that these four

factors significantly affect the decision metric, an appropriate parameter framework was designed, and a simulation example base was created. The example base contained over 900 training examples from nearly 300 simulations. The ID3 algorithm was used to determine the NSM Expert's classification "rules" in the form of *decision trees*. The performance of these decision trees was assessed on two arbitrary trajectories, and the performance results are presented using a predictive metric. The test trajectories used to evaluate the system's performance show that the NSM Expert adapts to new situations and provides reasonable estimates of the expected hybrid performance. The results also show how the NSM Expert chooses optimal or next-best navigation strategies when limited computational resources are available; in simple cases, its solutions are commensurate with the designer's intuition.

9. R. Stengel and L. Ryan, Stochastic Robustness of Linear Control Systems, *Proceedings of the 1989 Conference on Information Sciences and Systems*, The Johns Hopkins University, Baltimore, Mar 1989, pp. 556-561.**

A simple numerical procedure for estimating the *stochastic robustness* of a linear, time-invariant system is described. Monte Carlo evaluation of the system's eigenvalues allows the *probability of instability* and the related *stochastic root locus* to be estimated. This definition of robustness is an alternative to existing deterministic definitions that addresses both structured and unstructured parameter variations directly. This analysis approach treats not only Gaussian parameter uncertainties but non-Gaussian cases, including uncertain-but-bounded variations. Trivial extensions of the procedure admit alternate discriminants to be considered. Thus, the probabilities that stipulated degrees of instability will be exceeded or that closed-loop roots will leave desirable regions also can be estimated. Results are particularly amenable to graphical presentation.

10. R. Stengel and L. Ryan, Stochastic Robustness, oral presentation at the SIAM Conference on Control in the 90's, San Francisco, May 1989.**

This paper is an extension of the above paper with extended numerical results for the Doyle LQG counterexample. In particular, the comparison of stochastic robustness with singular-value analysis is pursued; whereas the latter provides only qualitative guidelines for "loop transfer recovery," the former provides a specific solution for the amount of robustness recovery required as a function of parameter uncertainty.

11. R. Stengel and L. Ryan, Multivariate Histograms for Analysis of Linear Control Systems, *Proceedings of the 1989 American Control Conference*, Pittsburgh, June 1989, pp. 937-943.**

A simple numerical procedure for estimating the *stochastic robustness* of a linear, time-invariant system is described. Based on Monte Carlo evaluation of the system's eigenvalues, this analysis approach introduces the *probability of instability* as a scalar measure of stability robustness. The related *stochastic root locus*, a portrayal of the root probability density, provides insight into robustness characteristics. Parameter variations are not limited to Gaussian parameter uncertainties; non-Gaussian cases, including uncertain-but-bounded variations, can be considered as well. Confidence intervals for the scalar probability of instability address computational issues inherent in Monte Carlo simulation. An example demonstrates stochastic robustness as applied to an aircraft control system in which parameters are alternately considered to have Gaussian, uniform, or binary probability distributions.

12. R. Stengel and L. Ryan, Application of Stochastic Robustness to Aircraft Control, *Proceedings of the 1989 AIAA Guidance, Navigation & Control Conference*, Boston, Aug 1989, pp. 698-708.**

Stochastic robustness, a simple numerical procedure for estimating the stability robustness of linear, time-invariant systems, is applied to a forward-swept-wing aircraft control system. Based on Monte Carlo evaluation of the system's closed-loop eigenvalues, this analysis approach introduces the *probability of instability* as a scalar stability robustness measure. The related *stochastic root locus* provides insight into robustness characteristics of the closed-loop system. Three linear-quadratic controllers of increasing robustness are chosen to demonstrate the use of stochastic robustness to analyze and compare control designs. Examples are presented illustrating the use of stochastic robustness analysis to address the effects of actuator dynamics and unmodeled dynamics on the stability robustness of the forward-swept-wing aircraft.

13. R. Stengel, Stochastic and Deterministic Analysis of Nonlinear Flight Control Systems, oral presentation at the *NASA Ames-UC Berkeley Nonlinear Flight Control Workshop*, Berkeley, Aug 1989.**

The flight dynamic equations for aircraft motion are inherently nonlinear, yet linear analysis has played a large role in the design and analysis of flight control systems. The principal reason is that the pilot has been

able to provide the underlying nonlinear trim control manually, and perturbations from the trim condition often are small. When large-amplitude perturbations from trim are important, nonlinear effects need to be considered. Deterministic models provide the basis for all analysis of aircraft motions. They are used directly to examine stability and response and indirectly to form the underpinnings of stochastic models. Stochastic effects fall into two distinct categories: uncertain response brought about by random inputs and initial conditions, and uncertain response arising from parametric uncertainty.

After briefly reviewing the nature of aircraft dynamics, a progression through four topics is made: *locally linearized control laws, adaptive nonlinear-inverse-dynamic (NID) control, analysis of control saturation effects, and stochastic robustness of closed-loop controllers*. Experiences with multi-input/multi-output linearized models are recounted, leading to validation of a linear-quadratic-Gaussian (LQG) pilot model by ground-based simulation and of a gain-scheduled LQG digital controller by helicopter flight test. Feedback linearization has been applied to the design of an all-axis controller for a general-aviation airplane, providing inherent protection against stalls and spins, as well as satisfactory flying qualities over the aircraft's entire flight envelope. This NID controller was shown to operate in real time using multiple microprocessors, and a mechanism for adapting it to the aerodynamic characteristics of individual aircraft was developed. Phase-plane analysis of the effects of control saturation on the closed-loop stability of open-loop-unstable aircraft is considered next, with stability boundaries defined by both open and closed regions of piecewise-linear-system trajectories. Finally, the topic of stochastic robustness is presented, based principally on results obtained for linear, time-invariant models of aircraft dynamics, with directions for application to nonlinear systems noted.

14. L. R. Ray and R. Stengel, Stochastic Stability and Performance Robustness of Linear Multivariable Systems, *Proceedings of the 1990 American Control Conference*, San Diego, May 1990.**

Stochastic robustness, a simple technique used to estimate the robustness of linear, time-invariant systems, is applied to a single-link robot arm control system. Concepts behind stochastic stability robustness are extended to systems with estimators and to stochastic performance robustness. Stochastic performance robustness measures based on classical

design specifications are introduced, and the relationship between stochastic robustness measures and control system design parameters is discussed. The application of stochastic performance robustness and the relationship between performance objectives and design parameters are demonstrated by means of an example. The results prove stochastic robustness to be a good overall robustness analysis method that can relate robustness characteristics to control system design parameters.

15. R. Stengel, Analysis of Stochastic Robustness in Linear Systems, oral presentation at the *Eighth Army Conference on Applied Mathematics and Computing*, Mathematical Sciences Institute, Cornell University, June 1990.**

This is an oral summary of material presented in earlier papers.

16. D. Linse and R. Stengel, System Identification for Nonlinear Control Using Neural Networks, *Proceedings of the 1990 Conference on Information Sciences and Systems*, Princeton University, Mar 1990.**

An approach to incorporating artificial neural networks in nonlinear, adaptive control systems is described. The controller contains three principal elements: a nonlinear-inverse-dynamic control law whose coefficients depend on a comprehensive model of the plant, a neural network that models system dynamics, and a state estimator whose outputs drive the control law and train the neural network. Attention is focused on the system-identification task, which combines an extended Kalman filter with generalized-spline function approximation (e.g., basis splines, back-propagation feedforward networks, or cerebellar model articulation controller networks). Continual learning is possible during normal operation (without taking the system off line for specialized training). Nonlinear-inverse-dynamic control requires smooth derivatives as well as function estimates, imposing stringent goals for the approximating technique.

17. D. Linse and R. Stengel, Neural Networks for Function Approximation in Nonlinear Control, *Proceedings of the 1990 American Control Conference*, San Diego, May 1990.**

Two neural-network architectures are compared with a classical spline interpolation technique for the approximation of functions useful in a nonlinear control system. A standard back-propagation feedforward neural

network and a cerebellar model articulation controller (CMAC) are presented, and their results are compared with a B-spline interpolation procedure that is updated using a recursive least-squares analysis. Each method is able to accurately represent a one-dimensional test function. Trade-offs between size requirements, speed of operation, and speed of learning indicate that neural networks may be practical for identification and adaptation in a nonlinear control environment.

ANNOTATED BIBLIOGRAPHY OF 1989-90 PUBLICATION

1. P. Shrivastava and R. Stengel, Stability Boundaries for Aircraft with Unstable Lateral-Directional Dynamics and Control Saturation, *J. Guidance, Control, and Dynamics*, Vol. 12, No. 1, Jan-Feb 1989, pp. 62-70.*

Aircraft that do not possess inherent stability must rely on closed-loop control systems for stable operation. Because there are limits on the deflections of an aircraft's control surfaces, the region of stable operation also is bounded. These boundaries are investigated for a lateral-directional example in which vertical fin size is inadequate to provide directional stability and where aileron and rudder deflections are subject to saturation. Fourth-order models are used for this study, with flight control logic based on minimum-control-energy linear-quadratic-regulator theory. It is found that the stability boundaries can be described by unstable limit cycles surrounding stable equilibrium points. Variations in regions of stability with gain levels and command inputs are illustrated. Current results suggest guidelines for permissible limits on the open-loop instability of an aircraft's lateral-directional modes.

Progress on Intelligent Guidance and Control for Wind Shear Encounter

D. Alexander Stratton
Princeton, University
Princeton, New Jersey

Low-altitude wind shear poses a serious threat to air safety. Avoiding severe wind shear challenges the ability of flight crews, as it involves assessing risk from uncertain evidence. A computerized intelligent cockpit aid can increase flight crew awareness of wind shear, improving avoidance decisions. This presentation outlines the primary functions of a cockpit advisory expert system for wind shear avoidance. It also introduces computational techniques being implemented to enable these primary functions.

WTA Wind Shear Probability Guidelines

The Wind Shear Training Aid contains a set of Microburst Wind Shear Probability Guidelines to assist flight crews in determining the risk of possible wind shear encounter. These guidelines relate various types of evidence that might be available in the cockpit to a discrete-valued "probability of wind shear encounter." When evaluating combinations of evidence, flight crews are instructed to combine the individual "probabilities" according to the rule, "LOW + MEDIUM = HIGH." If available evidence results in a "high probability of windshear encounter," the Wind Shear Training Aid states that a decision to avoid is appropriate. If the evidence results in a "medium probability of windshear encounter," then precautions are considered appropriate. The guidelines specify that the evidence must apply in the airport vicinity, during the intended time of operations, and along the low-altitude portion of the intended flight path. The Wind Shear Training Aid also supplies a set of Weather Evaluation Exercises to demonstrate the use of the guidelines. More importantly, the Wind Shear Training Aid states that the use of the guidelines "should not replace sound judgment in making avoidance decisions."

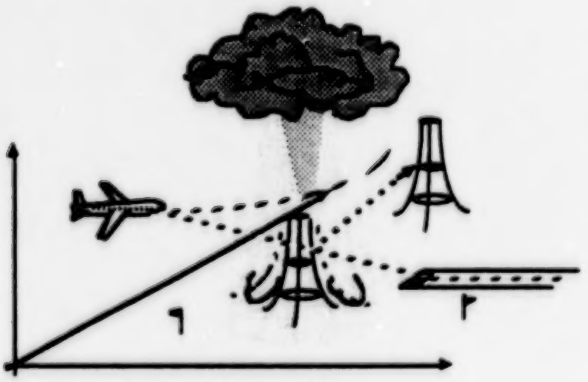
OBSERVATION	PROBABILITY OF WINDSHEAR
PRESENCE OF CONVECTIVE WEATHER NEAR INTENDED FLIGHT PATH:	
- With localized strong winds	HIGH
- With heavy precipitation	HIGH
- With rainshower	MEDIUM
- With lightning	MEDIUM
- With virga	MEDIUM
- With moderate or greater turbulence	MEDIUM
- With temperature/dew point spread between 30 and 50 degrees F	MEDIUM
ONBOARD WIND SHEAR DETECTION SYSTEM ALERT (Reported or observed)	HIGH
PIREP OF AIRSPEED LOSS OR GAIN:	
- 15 knots or greater	HIGH
- Less than 15 knots	MEDIUM
LLWAS ALERT/WIND VELOCITY CHANGE:	
-20 knots or greater	HIGH
- Less than 20 knots	MEDIUM
FORECAST OF CONVECTIVE WEATHER	LOW

Clues are cumulative: LOW + MEDIUM = HIGH

Decision Making for Wind Shear Avoidance

Wind shear is a spatial or temporal variation of the air about an aircraft that causes a deviation of the aircraft from its intended flight path. The hazard posed by severe low-altitude wind shear has prompted considerable study of the wind shear problem by meteorologists and aircraft engineers. A primary objective still being addressed is developing systems for improved flight crew capability to avoid severe wind shear.

A decision to avoid wind shear must be based on predictions made from available information. New and improved sensor systems under development could provide direct evidence of the presence of wind shear, but they will not be available in all circumstances. Wind shear avoidance must be based on indirect meteorological evidence when accurate sensors to determine the presence of wind shear are unavailable. A decision process based on a combination of meteorological evidence can provide better reliability than a decision process based on a single source.

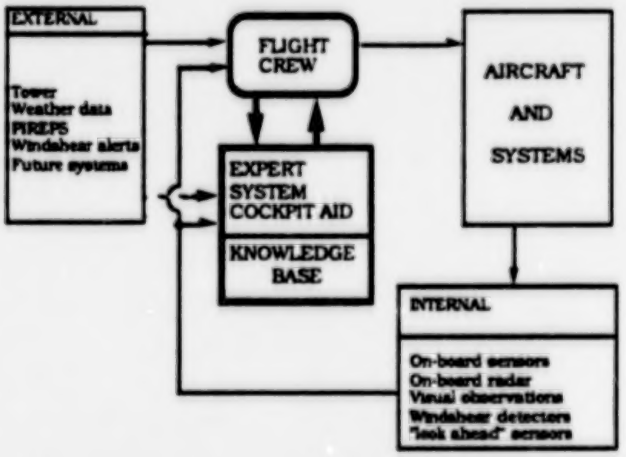


- Avoidance can enhance safety
- Avoidance involves prediction
  - Make predictions from current measurements
  - Interpret predictive information
- Avoidance involves uncertainty
  - Wind inputs must be inferred
  - Uncertain knowledge of wind shear meteorology
  - Indirect evidence can increase reliability

Intelligent Advisor for Wind Shear

Cockpit automation provides an opportunity to assist flight crews with decisions critical to avoiding low-altitude wind shear. The primary goals of a cockpit aid for wind shear avoidance are to increase the likelihood that the flight crew makes an avoidance decision when severe wind shear actually will be present and to decrease the likelihood of a decision to avoid when one is not called for. Technology from the rapidly-growing field of artificial intelligence provides a basis for a wind-shear-avoidance cockpit aid. In an intelligent system for decision aiding, the merits of possible decision alternatives are determined by a machine reasoning process that is rational and sound, producing results in an intuitive and meaningful manner. This means that the system is capable not only of recommending actions but of explaining why those actions are being recommended. An intelligent cockpit aid can summarize relevant information from a variety of sources and recommend actions, improving avoidance decisions.

An expert system for wind shear avoidance, dubbed the Wind Shear Safety Advisor, is depicted schematically on this slide. It will operate in real time, accepting evidence from on-board sensors and external evidence (such as pilot reports or PIREPs), perhaps facilitated by a direct data link (represented by a dotted line connecting the external sources to the expert system). It will interact with the flight crew, interpreting their intentions and providing advice and explanations. The development and improvement of the Wind Shear Safety Advisor are subjects of current research.



- Aid awareness, increase decision reliability
  - Summarize a variety of information
  - Meteorological knowledge
  - Real time prediction and estimation
  - Provide step-by-step explanations
  - Facilitated by direct data link

Wind Shear Safety Advisor

Primary Functions and Proposed Solution Methods

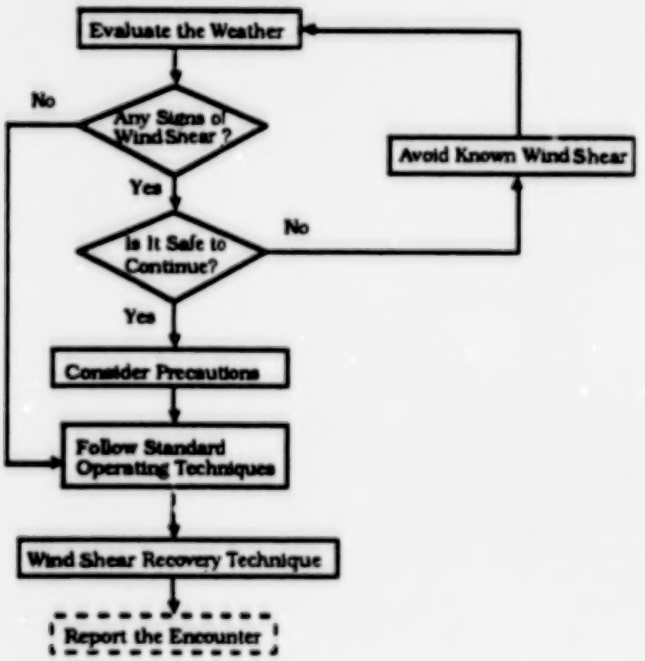
Primary functions of the Wind Shear Safety Advisor (WSSA) may be divided into the categories of executive, monitoring, prediction and assessment, and planning. Executive is a process of top-level control, while monitoring is the process of obtaining evidence from sources. Prediction is the process of extrapolating this evidence to the place and time of intended operations, while assessment is the process of considering the impact of evidence on the current flight plan. Planning is the process of altering the flight plan to reduce the risk of wind shear encounter. When operating in real time, these functions are executed in a cyclical fashion, processing evidence continuously. While the architecture currently being developed will facilitate real-time implementation, real-time implementation is a follow-on to the current study.

Declarative rule-based techniques and structured data base techniques have been used extensively in the development of the WSSA to provide a modular and explicit program structure. To deal with the uncertainty inherent to prediction and assessment, methods for probabilistic Bayesian inference are being incorporated. Algorithms for interpreting wind data based on stochastic flight path prediction are also being developed for risk assessment and planning.

- EXECUTIVE - Top-level control
  - *Rule-based techniques*
- MONITORING - Observe sensors, receive reports, alerts, warnings
  - *Rule-based techniques*
  - *Structured data bases*
- PREDICTION AND ASSESSMENT - Make predictions from reports, assess risk of continuing with nominal flight plan
  - *Probabilistic inference*
  - *Stochastic error prediction*
- PLANNING - consider flight path revisions and precautions, receive flight plan revisions
  - *Decision analysis*
  - *Rule-based techniques*
  - *Stochastic error prediction*

Wind Shear Training Aid
Model of Flight Crew Actions

The FAA Wind Shear Training Aid has been a primary source of knowledge for the Wind Shear Safety Advisor. The two-volume Wind Shear Training Aid document, prepared with the support of the Integrated FAA Wind Shear Program, was written by a team from the airframe industry that interacted with airlines, government, and academia. The Wind Shear Training Aid is a comprehensive training manual for flight crews that describes the hazards of wind shear and details precautionary and recovery procedures to help escape inadvertently-encountered wind shear.



Probability Theory for Risk Assessment

Developing a representation of these guidelines for the Wind Shear Safety Advisor that incorporates sound judgment is complicated by uncertainty surrounding wind shear meteorology and the subjective terms used in the guidelines. Use of the guidelines in real situations requires combined judgments involving uncertainty that the qualitative levels of probability and combination rule can not represent. Probability theory provides a set of sound and consistent axioms that are understood throughout the scientific community. Probability theory produces results consistent with ordinary logic and stochastic prediction and estimation. Bayesian belief network representation, which has been developed primarily by Judea Pearl [12], provides an efficient means to represent uncertain knowledge and to reason in a manner consistent with these axioms, provided some assumptions are made.

- F.A.A. Wind Shear Training Aid (1986)
  - Example training program for flight crews
  - Logical model of flight crew actions
  - Windshear avoidance guidelines
- Combined judgments involving uncertainty
  - "Convective weather near flight path"
  - "LOW," "MEDIUM," and "HIGH" Risk
  - "LOW + MEDIUM = HIGH"
- Probability and decision theory
  - Understood by scientific community
  - Relates to logic, stochastic estimation
  - Can incorporate statistical results
  - Efficient implementation using Bayesian network representation

Logical Expert System Techniques

Logical expert system techniques common to expert systems have been used for initial versions of the WSSA, as was previously reported. With rule-based techniques, knowledge pertinent to wind shear avoidance is transcribed as a set of statements of the form "If *premise* is true, then *consequent* follows." The expert system's data base is structured by defining a hierarchy of object structures, called *frames*. Structuring the data base adds modularity to rules and permits clearer explanations to be generated during execution.

Execution of the rule base is controlled by an *inference engine*. Forward- and backward- chaining searches, described in an earlier presentation, select rules to evaluate and update the data base through a process of logical deduction. As a side-effect of search, quantitative algorithms may be called as necessary. A recent implementation of the WSSA on a Symbolics 3670 LISP machine consists of over 200 rules.

- Declarative knowledge representation
  - "IF-THEN" rules
  - Symbolic parameters, LISP functions
  - Data base structured with frames
- Inference engine controls rule evaluation
  - Forward- and backward- chaining
  - Quantitative algorithms called conditionally as side-effects of search
- Symbolics 3670 LISP Machine
  - 212 "IF-THEN" rules
  - 78 numerical and symbolic parameters
  - 15 frame structures
  - 80 functions (& Common LISP)

Elements of Probabilistic Reasoning

In a probabilistic model of reasoning, such as a Bayesian network, reasoning is a process of conditioning the probabilities of hypotheses on the evidence. Bayes's rule is used to accomplish this. To illustrate the fundamentals of probabilistic reasoning, let H represent a hypothesis that wind shear is on the intended flight path, and let E1 represent a piece of evidence, such as a wind shear detection system alert. From a prior probability that wind shear is on the intended flight path, Pr{H}, the posterior probability is computed using Bayes's rule. To use Bayes's rule, we must provide the probability of detection of the detection system, Pr{E1 | H}. The prior probability of receiving the evidence, Pr{E1}, must also be specified; however, it is usually easier to provide Pr{E1 | ¬ H}, the probability of false alarm (the symbol ¬ denotes logical negation), and compute Pr{E1}. These equations provide a basis for more general probabilistic reasoning, but additional assumptions must first be made.

To see why this is so, suppose that a second detection system alert, E2, is received. Use of Bayes's rule requires the computation of Pr{E1, E2 | H}, a joint probability distribution. In general, the computation of joint probability distributions is a cumbersome process. To simplify the process, we assume that wind shear is the cause of the alerts and that the alerting systems operate independently of each other. Thus, if wind shear is present, the probability of receiving the second alert is not changed by the arrival of the first alert. This assumption is known as *conditional independence*, and we say that the effects are *conditionally independent given the cause*. The assumption of conditional independence simplifies Bayes's rule, permitting efficient computation of the posterior probability of H, based on the probability of detection and false alarm rate of the individual systems.

- Condition probabilities of hypotheses on evidence: Bayes's rule

  Hypotheses of wind shear (H)

  Reports of wind shear (E1 and E2)

  $$Pr\{H \mid E1,E2\} = Pr\{E1,E2 \mid H\}Pr\{H\}/Pr\{E1,E2\}$$

- Conditional independence assumption: reports independent consequents of wind shear

  $$Pr\{H \mid E1,E2\} = Pr\{E2 \mid H\}Pr\{H \mid E1\}/Pr\{E2 \mid E1\}$$

- Required statistics:

  Prior probability of wind shear, Pr{H}

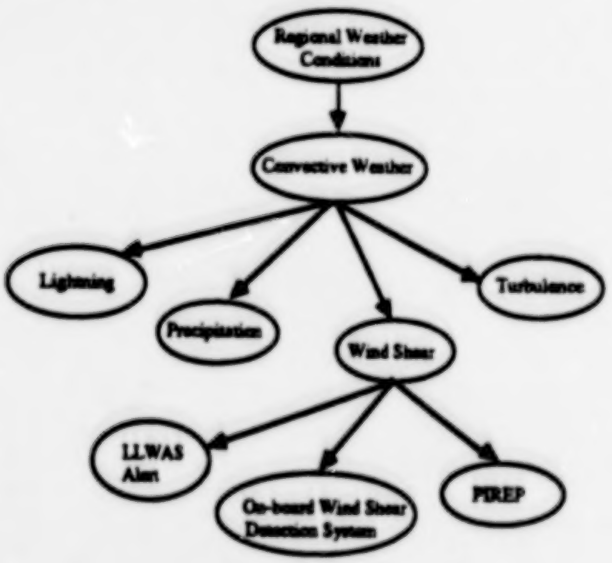
  Probability of detection, Pr{E1 | H}

  Probability of false alarm, Pr{E1 | ¬H}

Bayesian Network Interpretation of Wind Shear Probability Guidelines

A Bayesian network is a probabilistic model of a system. Bayesian networks use graphical representations of dependency, where a set of discrete random variables are represented as nodes and where uncertain relationships between the variables are represented as links. Cause-effect relationships are represented graphically by adding arrows to the links, pointing from cause to effect. A graphical model of dependencies surrounding the meteorology of wind shear, constructed from the Microburst Wind Shear Probability Guidelines and extended to include the prediction, is shown here. Elements of the guidelines are represented as random variables, which can take one of a set of exhaustive and mutually-exclusive values.

Network representations enable efficient probabilistic reasoning because all of the dependencies between variables are specified by the links. In Bayesian networks, nodes that are "effects" or manifestations of the same cause are assumed to be conditionally independent given the cause. With this assumption, all of the information necessary to condition the probability distribution at any particular node can be obtained from the nodes to which it is directly linked. If reasoning is simulated by sequential processing, the probability distribution at each node can be updated step by step, permitting explanations to be generated at each step. Alternatively, the network could be implemented on a set of identical processors, permitting several steps to occur concurrently.



Graphical representation of dependency

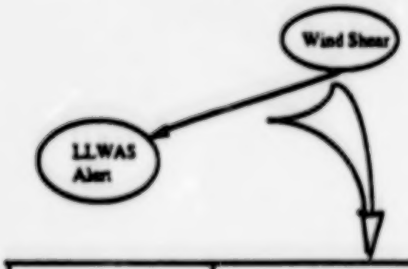
Probabilities stored at each node

Nodes updated using Bayes's rule, axioms

Requirements for Probabilistic Representation

The construction of a Bayesian network begins with the construction of a graphical model of dependency, such as the previous slide. After constructing the graphical representation, and determining a set of values for each variable, the knowledge base is augmented by the definition of conditional probabilities associated with each link. For example, completing the link between node Wind Shear and node LLWAS Alert (which represents the state of the Low Level Wind Shear Alert System) requires the estimation of the probabilities shown. Reliability statistics for this link were derived from data obtained for the enhanced LLWAS and incorporated into the network. For each of the other links of the network, a set of conditional probabilities must be defined. These sets of probabilities are called link matrices. The network definition is completed by the definition of prior probability distributions at every node in the network.

There are no statistics available for some of the wind shear link matrices, such as the link between Precipitation and Convective Weather, indicating specific areas where additional meteorological research is required. Definition of these probabilities is aided by qualitative information in the Wind Shear Training Aid and other sources, but the actual values must be assigned subjectively at this time. This is less than desirable, for it adds a degree of subjectivity to some of the probabilities generated by the system during reasoning. Nevertheless, the subjectivity of the values can be reduced by the acquisition of more knowledge. Subjective language contained in the Microburst Wind Shear Probability Guidelines and other sources invites the developer to consider more detailed definitions for the concepts, guiding the search for additional meteorological knowledge and statistical data.



LLWAS Alert (LA)	Wind Shear (WS)	
	[illegible] on flight path (WS1)	[illegible] on flight path (WS2)
LLWAS Alert of Severe Wind Shear (LA1)	Pr(LA=LA1 | WS=WS1) = 0.714	Pr(LA=LA1 | WS=WS2) = 0.027
LLWAS Alert of Moderate Wind Shear (LA2)	Pr(LA=LA2 | WS=WS1) = 0.095	Pr(LA=LA2 | WS=WS2) = 0.554
No LLWAS Alert (LA3	Pr(LA=LA3 | WS=WS1) = 0.191	Pr(LA=LA3 | WS=WS2) = 0.419

Conditional and prior probabilities

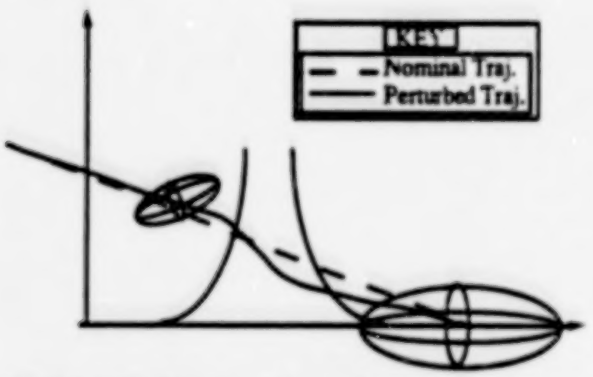
Statistical data incorporated

Some values assigned subjectively

Flight Path Error Prediction With Wind Statistics

If direct measurements of the winds are available, wind shears that are truly hazardous to the aircraft can be distinguished from less hazardous shears. A reasonable approach to distinguish hazardous wind shears using wind measurements is based on flight path deviations from the aircraft's nominal flight path. A hazardous wind shear can be distinguished by determining whether or not a perturbed trajectory lies outside a critical region in space. A perturbation model of the aircraft and control law provides an efficient means to compute trajectory deviations.

Providing a means of distinguishing hazardous wind shear is complicated by errors in wind measurements. For example, devices based on the doppler effect can detect only the radial component of winds; vertical winds that can affect the aircraft's flight path cannot be detected. Stochastic prediction techniques provide a means of computing stochastic trajectory errors, such as touchdown dispersion or mean square glide slope error, given uncertain but statistically defined data from wind sensors.



- Flight path deviations through wind shear
  - Wind profile, nominal trajectory, control law
  - Perturbation model, control saturation
- Covariance propagation
  - Error statistics of wind data
  - Correlation of horizontal and vertical winds
  - Correlation with meteorological parameters
- Stochastic error function computation
  - Touchdown error, impact velocity
  - Errors in glideslope, altitude, energy rate

Conclusions

Cockpit automation provides the opportunity to aid flight crew decisions relevant to wind shear and improve flight safety, but the decision-making process involves the management of uncertainty. Bayesian belief network techniques provide a means to represent this subjective meteorological knowledge so that the knowledge can be used in cooperation with more accurate wind shear detection systems. With the probabilistic representation of the Microburst Wind Shear Probability Guidelines, the door is open for the incorporation of a wider variety of meteorological expert knowledge. This means of representation also facilitates the incorporation of more reliable wind shear detection systems.

Stochastic prediction algorithms provide a means of interpreting uncertain but statistically-defined wind data. An error prediction algorithm would incorporate error models of wind sensors. Stochastic prediction of a trajectory error can provide the basis for an avoidance decision, once a suitable error metric is defined.

- Bayesian networks provide a means to interpret Wind Shear Training Aid avoidance guidelines
  - Based on probability theory
  - Enables incorporation of statistical data
  - Suggests additional meteorological studies
- Stochastic prediction techniques provide a means to incorporate wind data
  - Base decisions on error analysis
  - Requires error model of wind sensors
  - Requires definition of error metric

Flight Simulation for Wind Shear Encounter

Sandeep S. Mulgund
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey

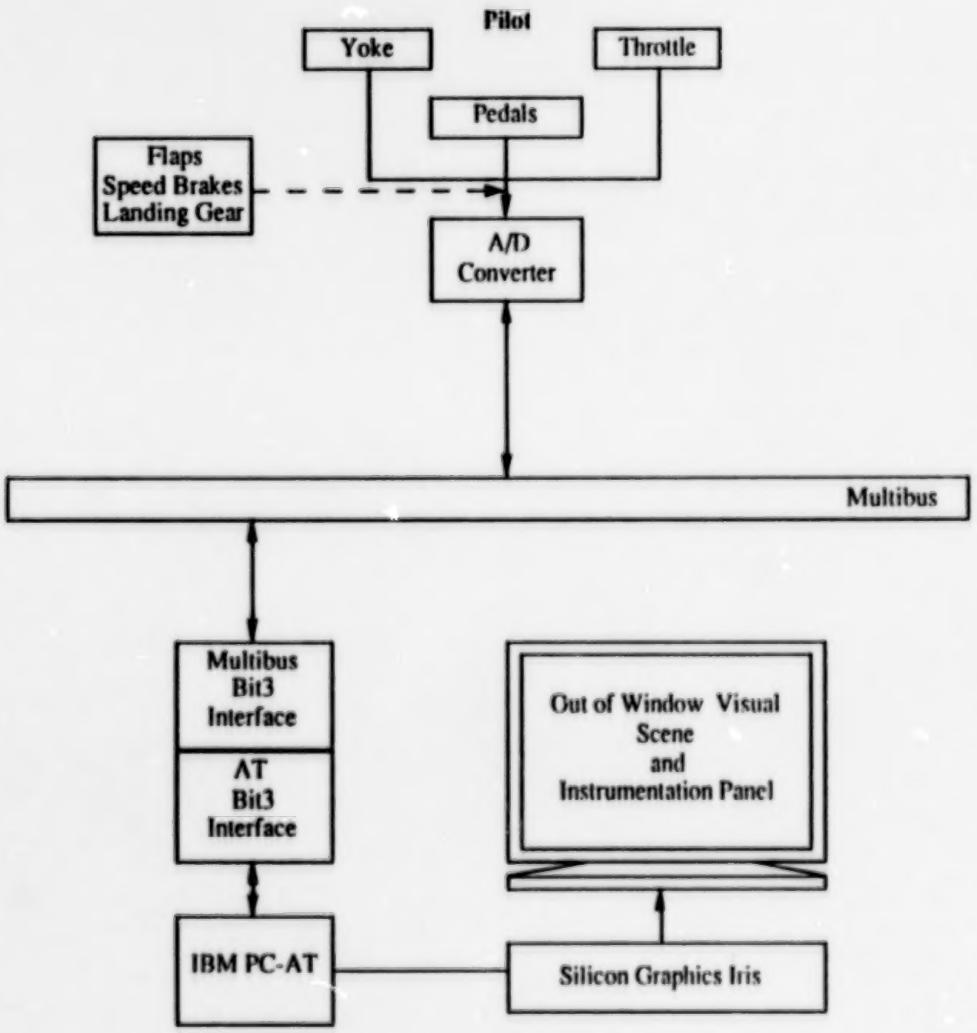
Introduction

A real-time piloted flight simulator is under development in the Laboratory for Control and Automation at Princeton University. This facility will be used to study piloted flight through a simulated wind shear. It will also provide a testbed for real-time flight guidance laws. The hardware configuration and aerodynamic model used are discussed. The microburst model to be incorporated into the simulation is introduced, and some proposed cockpit display concepts are described.

Simulator Architecture

The simulator presently consists of two computer subsystems. An IBM PC-AT contains the aerodynamic model of the aircraft. This computer samples pilot inputs through a 12 bit A/D board and integrates the aircraft equations of motion. Aircraft attitude and position variables are transmitted by the AT to a Silicon Graphics IRIS Workstation, which in turn generates an external visual scene and a primary flight instrument display. The present visual scene contains a representation of Teterboro Airport and a simplified depiction of the Manhattan skyline. The flight instruments are formatted to resemble a modern electronic flight instrumentation system (EFIS). The display is partitioned so that the top half of the screen contains the out-of-window display, while the bottom half displays the EFIS. A graphical representation of a Boeing 737 instrument panel has also been developed on the IRIS.

LCA Flight Simulator Hardware Architecture



Simulator Aerodynamic Model

The simulator is presently configured to represent the dynamics of a Boeing 737 Transport. The aerodynamic data for the aircraft are taken from the Sperry TCV Simulation developed for NASA Langley. A major portion of these data comes from wind tunnel testing, corrected for Reynold's number and aeroelastic effects. Much of it has been confirmed in flight testing of the 737, and adjusted if necessary to match actual 737 flight characteristics. As such, the Sperry simulation provides a faithful representation of 737 dynamics over a broad flight envelope. Nonlinearities such as stall behavior and control saturation are modeled in the simulation. These effects may be significant in a wind shear environment, where high angles of attack and full control capacity could be required to escape a hazardous situation.

- Nonlinear model of aircraft dynamics developed by Sperry Systems for NASA Langley

- Simulation based on the 6 DOF force and moment equations of the airplane

- Solution of these equations yields the flight path and angular orientation of the airplane at each instant in time

- Aerodynamic loads on aircraft are functions of

    - Mach number
    - Altitude
    - Incidence angles
    - Rotation rates
    - Control deflections
    - Geometry changes (flap, gear, & spoiler deflection)
    - Ground proximity

Analytic Downburst Model

A microburst model has not yet been interfaced with the aircraft equations of motion. The present plans are to use an analytic model developed by Rosa Oseguera and Rowland Bowles at NASA Langley. This simple time-invariant model permits the simulation of different shears through the specification of four parameters. Wind components and spatial derivatives are easily obtained in cartesian coordinates from the velocity equations. This downburst model will be integrated with the equations of motion of the aircraft, and used to study conceptual wind shear hazard and flight guidance displays.

- Developed at NASA Langley by Oseguera and Bowles

- Represents an axisymmetric stagnation point flow, based on velocity profiles from the Terminal Area Simulation System model

- Time-invariant model

- Altitude dependence includes boundary layer effects near the ground, and closely matches real-world measurements

- Permits simulation of different shears through specification of four parameters:

    1) radius of the downdraft column
    2) maximum wind velocity (horizontal or vertical)
    3) altitude of maximum outflow
    4) depth of outflow

- Wind components in x,y, and z directions are easily obtained from velocity equations for a given aircraft position relative to downburst core

Display of Wind Shear Information

Once the microburst has been incorporated into the simulation, it will be necessary to augment the flight displays in some manner to provide the pilot with some indication of the potential hazard due to the wind shear. The type of display that will be considered assumes some sort of forward-look wind shear information. The hazard to the aircraft will be characterized by the "F-Factor", a concept developed at NASA Langley. This parameter completely describes the effect of a wind shear on the performance of an aircraft, in terms of its constant speed climb gradient capability. A persistent value of F greater than about 0.15 will exceed the climb capability of most transport aircraft.

The display concept under consideration will graphically present the anticipated F-Factor ahead of the aircraft, possibly in a head-up format. A display concept has been developed at NASA Langley that uses anticipated F-Factor to predict the aircraft flight path. Such a display may be developed for the simulator at Princeton, but initially the F-Factor will be used to characterize the wind shear hazard.

- Characterize the intensity of the wind shear using the "F-Factor":

$$F = \frac{\dot{W}_x}{g} + \frac{W_h}{V}$$

where

$$\dot{W}_x = \frac{\partial W_x}{\partial x}\dot{x} + \frac{\partial W_x}{\partial h}\dot{h} + \frac{\partial W_x}{\partial t}$$

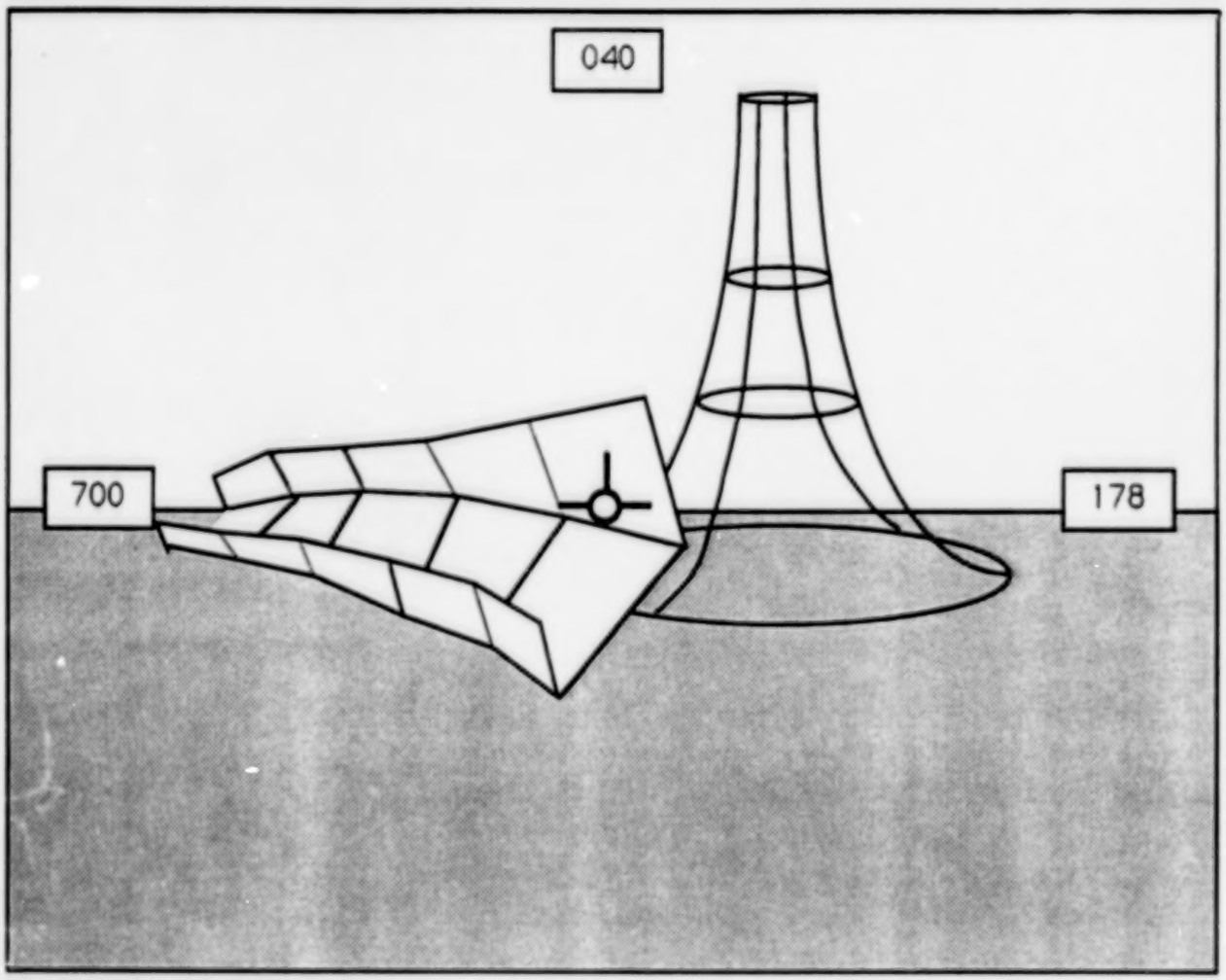
- F is a function of the aircraft's trajectory, and relates directly to its constant airspeed climb gradient capability

- Objective is to visually represent the F-Factor ahead of the airplane, to characterize the anticipated hazard due to the wind shear

- Flight guidance commands must also be presented to the pilot

"Highway in the Sky" Display Concept

In addition to a hazard display, it would be useful to have some sort of flight guidance command display. Typically, this takes the form of a set of command bars on the artificial horizon. A possible alternative to this conventional format is the so called "Highway in the Sky". Developed at the Flight Dynamics Laboratory at Wright Patterson Air Force Base, this display predicts the future aircraft trajectory based on the current state and the guidance strategy generated by the flight management computers. Rather than predicting future aircraft states, it would be interesting to use such a display to plot an escape profile out of a microburst. This would require the existence of an escape strategy, which could include lateral maneuvers. The "highway in the sky" would then display the flight path the pilot should take to avoid the wind shear hazard.



Future Work

- Implement the Oseguera-Bowles wind shear model in the aircraft's equations of motion

- Develop displays to characterize wind shear hazard, and investigate alternatives to "command bars"

- Continue development of simulator hardware

STOCHASTIC ROBUSTNESS: TOWARDS A COMPREHENSIVE ROBUSTNESS TOOL

Laurie Ryan Ray

Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey

Stochastic robustness is a simple technique to determine the robustness of linear, time-invariant systems by Monte Carlo methods. *Stochastic stability robustness* has been described previously. These results are extended here to provide insight into control system design for performance. Together, stochastic stability and performance robustness concepts constitute a comprehensive tool that can be used to analyze control system robustness properties. As well, they offer control system design insight that can set the stage for stochastic robustness synthesis. The concept of stochastic stability robustness is reviewed, stochastic performance robustness is introduced, and stochastic robustness synthesis is described qualitatively. Confidence intervals necessary for comparing control laws statistically are presented.

STOCHASTIC ROBUSTNESS ANALYSIS

Stability robustness

Performance robustness

STOCHASTIC ROBUSTNESS SYNTHESIS

Design insight uncovered by stochastic robustness

Confidence intervals for comparing control laws

SUMMARY

STOCHASTIC ROBUSTNESS ANALYSIS

Stochastic robustness is a robustness measure based on the probability of satisfactory stability/performance in the face of uncertainty. Stability robustness is described by a single metric: the probability of instability. Because it is a statistical measure of robustness, and because it directly uses knowledge of the statistics of the physical parameter variations, stochastic robustness is inherently intuitive and precise. The physical meaning behind the probability of instability is apparent, and overconservative or insufficiently robust designs can be avoided. Concepts behind stochastic stability robustness can be easily extended to provide insight into control system design for performance. Design specifications such as rise time, overshoot, settling time, dead time, and steady-state error are normally used as indicators of adequate performance and lend themselves to the same kind of analysis as described above. Concepts of stochastic stability robustness analysis can be applied to these criteria giving probabilistic bounds on individual scalar performance criteria. Stochastic robustness concepts can be applied to specific aircraft handling qualities criteria as well. Binomial confidence intervals for the scalar probability of instability have been presented, and these apply to performance robustness criteria as well.

STOCHASTIC ROBUSTNESS

A robustness *measure* based on the
probability of satisfactory stability/performance,
given the statistics of a plant's parameter variations

STABILITY ROBUSTNESS: PROBABILITY OF INSTABILITY

PERFORMANCE ROBUSTNESS: MANY MEASURES

Degree of stability
Time response envelopes
Handling qualities criteria

BINOMIAL CONFIDENCE INTERVALS

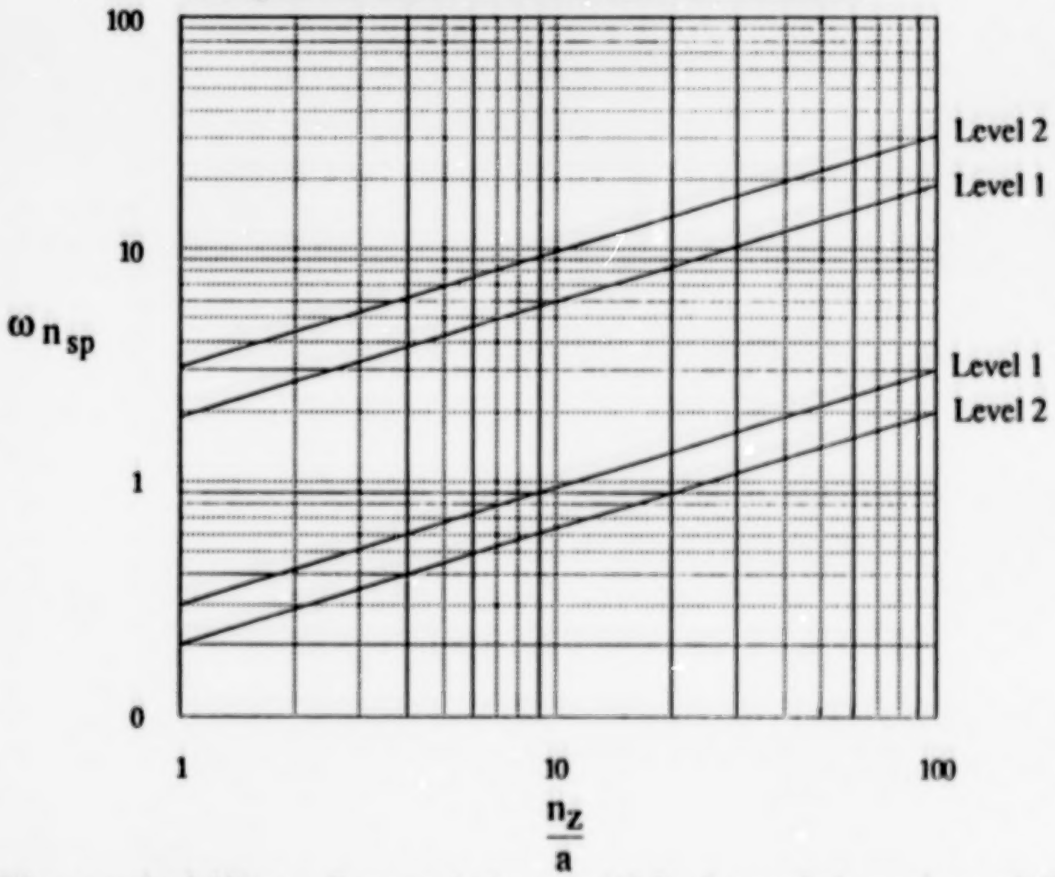
Statistical bounds for the
estimated stability or performance measure

HANDLING QUALITIES ROBUSTNESS

Principles behind stochastic stability robustness can be directly applied to aircraft handling qualities. Here, the short-period mode is evaluated using the MIL-F-8785C specification that relates short-period handling quality levels to the normal acceleration sensitivity to angle-of-attack vs. short-period undamped natural frequency. Each Monte Carlo evaluation would result in a single point on the graph. The probability of remaining within level 1 or 2 specifications is the performance robustness metric. The abscissa and ordinate quantities can be computed using very little computation beyond eigenvalue evaluation. Hence, performance can be characterized as easily as stability using this metric.

Principles behind stochastic stability robustness can be directly applied to aircraft handling qualities

Example: Short-Period Frequency Requirements from MIL-F-8785C

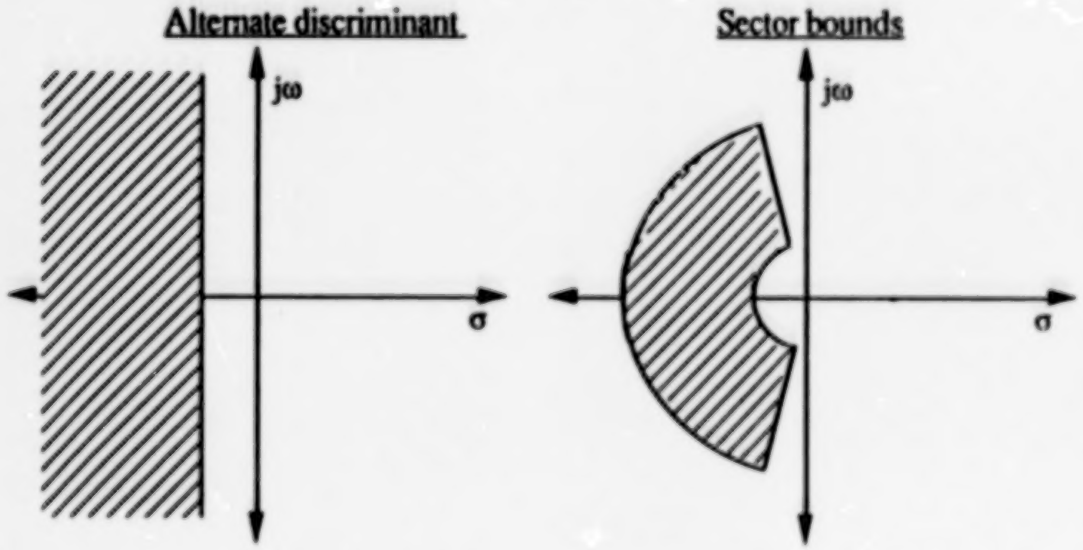


The probability of remaining within Level 1 or Level 2 regions is the scalar performance robustness metric

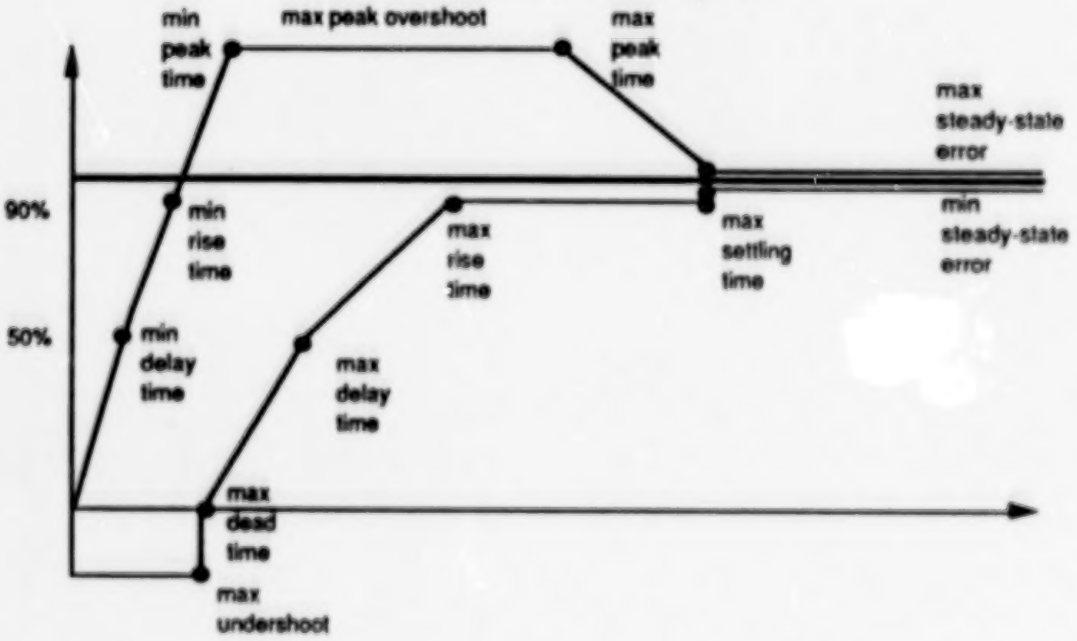
PERFORMANCE ROBUSTNESS

Making the transition from strict stability to degree of stability is simple. Alternate regions can be described that relate to classical measures of response speed. Alternate discriminants relate to time-to-half or time-to-double. Sector bounds relate directly to damping ratios and natural frequencies. The probability of closed-loop eigenvalues remaining within the alternate region is the scalar performance robustness metric. Time response envelopes can be defined as well, and stochastic robustness analysis gives the probability of a response remaining within the desired envelope. Using any of these metrics, binomial confidence intervals apply.

Transition from strict stability to degree of stability



Time response envelopes

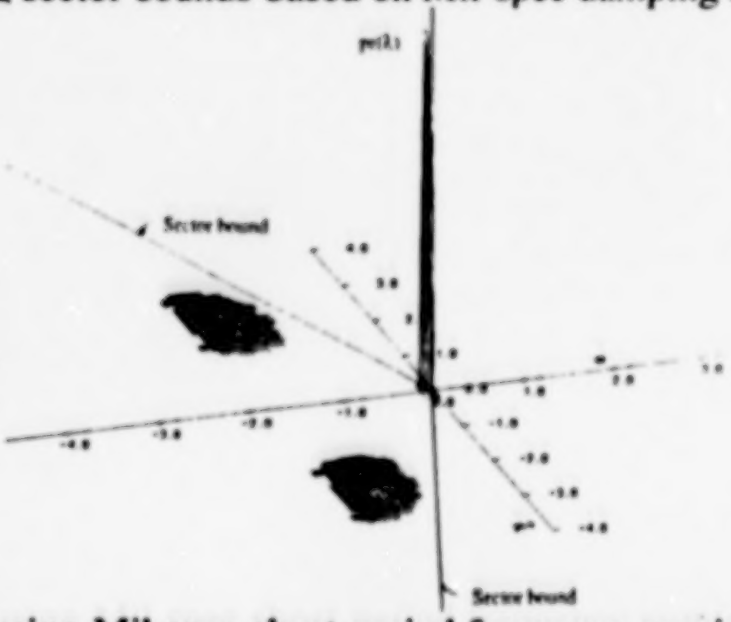


TWIN-JET TRANSPORT AIRCRAFT: SHORT-PERIOD MODE EVALUATION

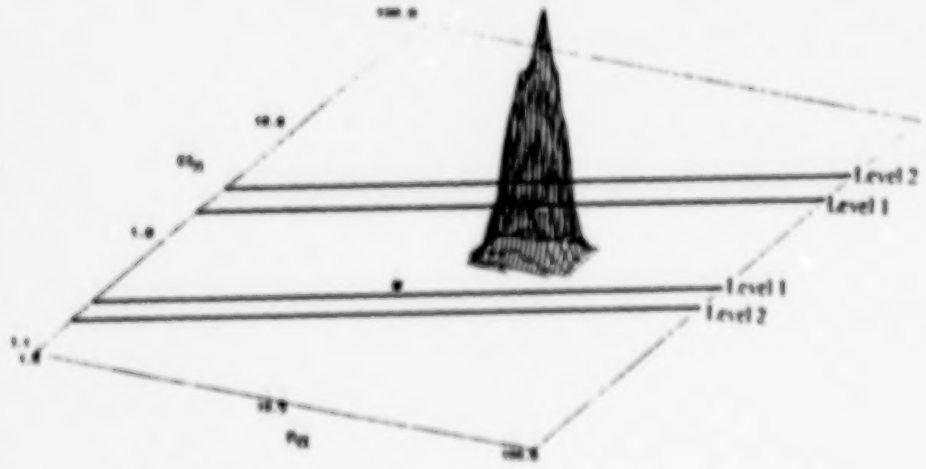
As an example of the above concepts, stochastic performance robustness analysis is applied to a nonlinear longitudinal model of a twin-jet transport aircraft. Each Monte Carlo evaluation consists of linearizing the system around nominal trim conditions and computing the eigenvalues and $\frac{n_z}{\alpha}$. The probability of violating sector bounds and the probability of violating level 1 Mil-spec requirements are shown here.

Nonlinear, longitudinal rigid-body model, 22 parameters

..using sector bounds based on Mil-spec damping requirement



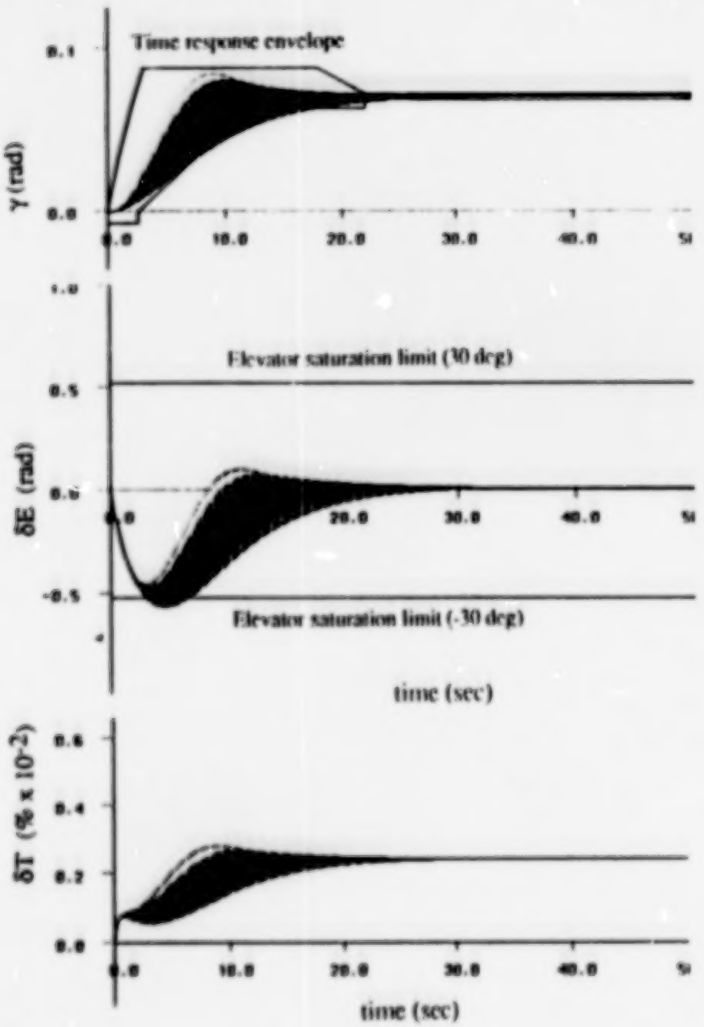
..using Mil-spec short-period frequency requirement



TWIN-JET TRANSPORT AIRCRAFT: CLOSED-LOOP COMMAND RESPONSE

Closed-loop flight-path angle command response is given here, where the controller was developed using a Proportional-Filter-Implicit-Model-Following LQR control law. Here, the probability of violating the defined time response envelope and the probability of control saturation are the performance robustness metrics.

Flight-path-angle command response distribution



$\hat{P}_{\gamma} = 0.034$ (0.0199,0.0539)

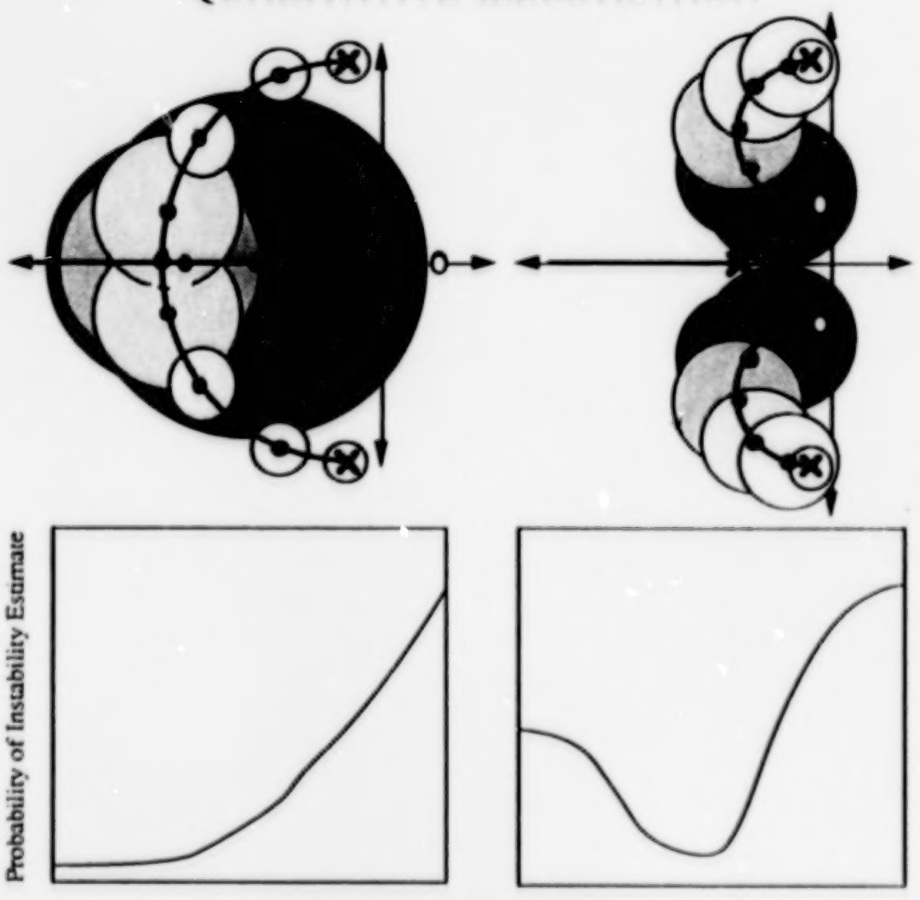
$\hat{P}_{\delta E} = 0.502$ (0.457, 0.547).

STOCHASTIC ROBUSTNESS SYNTHESIS

Metrics resulting from stability and performance robustness can be related to controller parameters, providing a foundation for design tradeoffs and optimization. This is illustrated qualitatively here. As gain increases along a root locus, uncertainty is magnified, and the "uncertainty circles" associated with the closed-loop root locations enlarge. The first root locus demonstrates one where the probability of instability increases monotonically with increased gain. The shape of the second root locus makes it possible for the uncertainty circles to initially cross into the right-half plane, remain in the left-half plane as gain increases, and finally cross into the right-half plane again for very large gain. The probability of instability decreases, then increases, and it has a minimum for some value of gain.

Tradeoffs exist and can provide a foundation for optimization

QUALITATIVE ILLUSTRATION



Closed-loop roots are enclosed by "uncertainty circles"
a) Stability robustness decreases as gain increases
b) Stability robustness increases, then decreases

STATISTICAL COMPARISON OF CONTROL LAWS

The above analysis qualitatively demonstrates tradeoffs that can provide a foundation for optimization. The ability to statistically compare control laws and say with certainty that one is better than another is another tool necessary for optimization. The Bonferroni inequality can be used to describe confidence intervals for the difference based on individual confidence intervals. As illustrated, the one control law is statistically better than another when their individual confidence intervals no longer overlap.

Given binomial confidence intervals for two estimates

$$\Pr(L_1 \leq \mathbb{P}_1 \leq U_1) = 1 - \alpha_1$$
$$\Pr(L_2 \leq \mathbb{P}_2 \leq U_2) = 1 - \alpha_2$$

the confidence interval for $\Delta\mathbb{P} = \mathbb{P}_1 - \mathbb{P}_2$

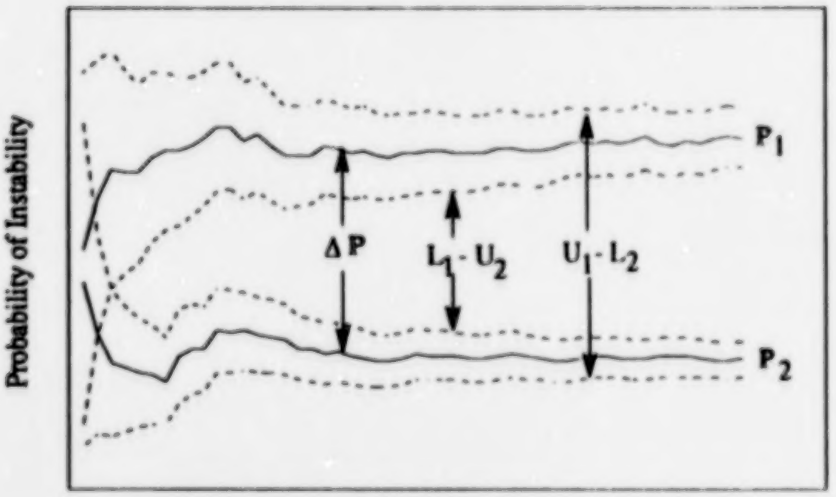
$$\Pr[(L_1 - U_2) \leq \Delta\mathbb{P} \leq (U_1 - L_2)] \geq 1 - \alpha$$

is derived from the Bonferroni inequality

$\alpha = \alpha_1 - \alpha_2$ for dependent intervals

$\alpha = \alpha_1 - \alpha_2 + \alpha_1 \alpha_2$ for independent intervals

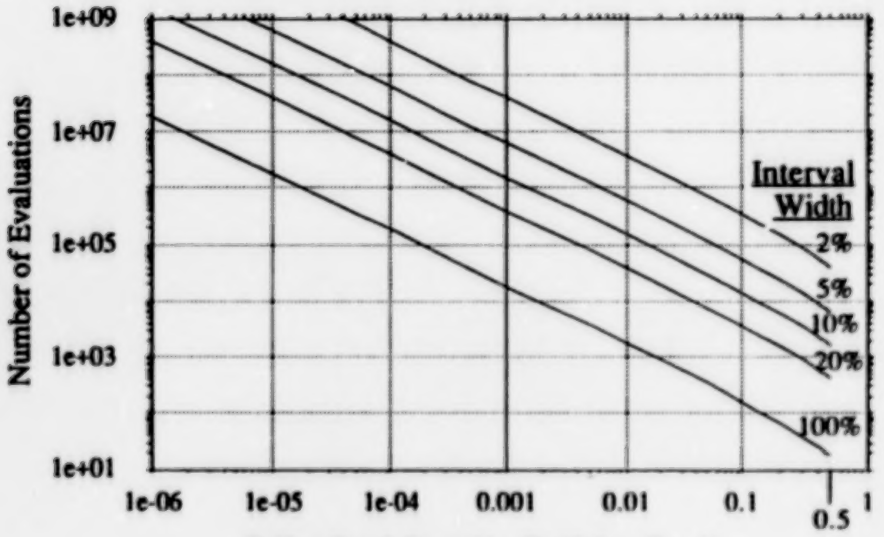
ILLUSTRATION



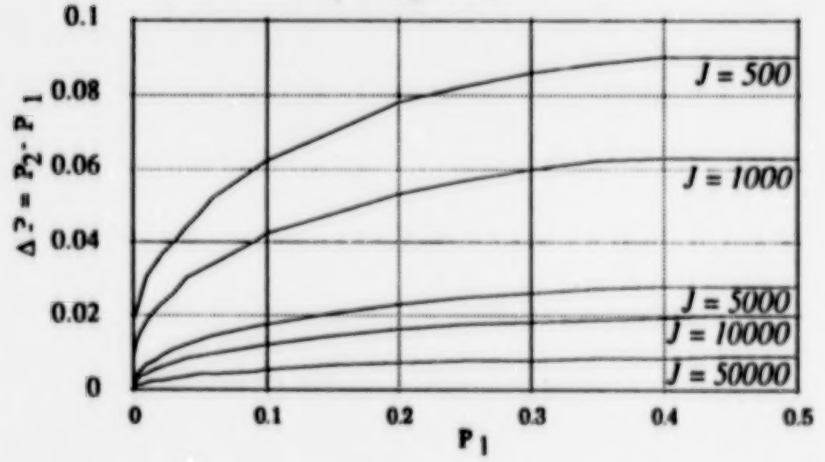
HOW MANY EVALUATIONS ?

The number of evaluations necessary to distinguish between two control laws is an important factor when considering stochastic robustness synthesis using optimization. Here, the number of evaluations necessary for a required interval width is related to the number of evaluations necessary to distinguish a significant difference.

Number of Evaluations
for a Required Interval Width and $\alpha = 0.05$



Number of Evaluations Required
to Identify a Significant Difference
$\alpha_1 = \alpha_2 = 0.05$



SUMMARY

Stochastic robustness is a good overall robustness analysis tool. It is intuitive and simple, and makes good use of engineering knowledg. Both stability and performance metrics can be identified, and confidence intervals offer statistical significance to the resulting metrics. It is a good candidate for synthesis techiques as well, because it demonstrates robustness tradeoffs with control systems design parameters. Confidence intervals for the difference between two probabilities provide a tool that can be used in future optimization studies.

Good overall robustness analysis tool

Intuitive and simple

Makes use of engineering knowledge

Both stability and performance robustness metrics

Confidence intervals are easily interpreted

Good candidate for control system synthesis

Shows stability and performance tradeoffs with design parameters

Confidence intervals for differences can be defined

Neural Networks in Nonlinear Aircraft Control

Dennis J. Linse
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey

Recent research indicates that *Artificial Neural Networks* offer interesting learning or adaptive capabilities. The current research focuses on the potential for application of neural networks in a nonlinear aircraft control law. The current work has been to determine which networks are suitable for such an application and how they will fit into a nonlinear control law.

Parameter Estimation in Nonlinear Control

The equations of motion of an aircraft can be cast into a set of nonlinear ordinary differential equations that are linear in the control input, **u**. Included in the state and output equations is a parameter vector, **p**. Several methods, including Nonlinear Inverse Dynamics [Lane and Stengel, 1988] and decoupling [Singh and Rugh, 1972], are available for generating nonlinear feedback laws if the system is known. Using these techniques, the feedback laws are determined as functions of the parameter vector, **p**. **p** itself is a function of the current state and possibly external states and controls not included in the system dynamics. The main difficulty is to provide an estimate of the parameters that are possible complex nonlinear functions of the states.

Given a Nonlinear Dynamic System (e.g. an aircraft)

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{p}) + \mathbf{G}(\mathbf{x}, \mathbf{p})\mathbf{u}$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{p})$$

and a Nonlinear Feedback Law

$$\mathbf{u} = \mathbf{a}(\mathbf{x}, \mathbf{p}) + \mathbf{B}(\mathbf{x}, \mathbf{p})\mathbf{v}$$

where

$$\mathbf{p} = \mathbf{p}(\mathbf{x}, \mathbf{x}_{ext}, \mathbf{u}_{ext})$$

how do you estimate the parameters, **p** ?

Nonlinear Function Approximation

While many techniques exist for nonlinear function approximation, three have been chosen for further investigation. A B-spline (basis spline) technique with coefficients updated using Recursive Least-Squares estimation represents a classic function approximation method. Two neural networks methods are also investigated. The Back-Propagation Feedforward Network [Rumelhart, Hinton, and Williams, 1986] is a popular, widely investigated, model in the neural network community. The Cerebellar Model Articulation Controller (CMAC) [Albus, 1975] is less well known, but very useful and powerful in function approximation implementations.

Three appealing methods

- Classic method
  - Recursive Least Squares Estimation using B-Splines
- Neural methods
  - Back-Propagation Feedforward Network
  - Cerebellar Model Articulation Controller (CMAC) Network

Comments on Operation and Learning

Each of the three techniques has different adaptation (or learning) capabilities. The B-spline method requires a matrix inversion that may be very slow, especially for high input dimension systems. The least-squares estimation scheme effectively extracts information from the input data, making adaptation fast when measured in terms of number of points presented to the estimator. The feedforward network, on the other hand, needs many presentations to accurately approximate the nonlinear function. Once trained, it can be extremely fast in operation, especially if fully implemented on a VLSI chip. The CMAC quickly approximates the desired function in the neighborhood of the training points and provides good operation speed on a traditional computer architecture.

- Recursive Least Squares B-Splines
  - Slow learning
    Matrix inversion
  - Effective learning
- Back-Propagation Feedforward Networks
  - Relatively Inefficient learning
    Many training points needed
  - *FAST* operation
- CMAC
  - Reasonable learning and operation

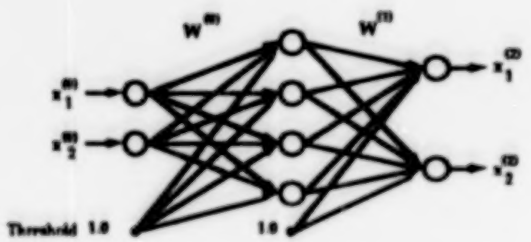
Approximation by Neural Networks

The approximation capabilities of traditional spline techniques are relatively well-known. The capabilities of the neural networks are less well-known. The CMAC approximates by a generalized table look-up. Using overlapping, quantized inputs, the CMAC output is a piece-wise continuous approximation of the input function. The approximation accuracy is limited by the size of the table used in the look-up scheme. The feedforward network provides a continuous approximation of the desired function with the accuracy determined by the number of layers and nodes in the network architecture. There has been much recent interest in exactly determining the approximation capabilities of such networks.

- CMAC's approximate by generalized table look-up
  - Approximation limited by table size
  - Stair-step output
- Feedforward networks
  - Capacity determined by nodes and layers
  - Continuous output
  - Much recent interest in approximation abilities

Back Propagation Neural Networks

A very simple, 2–input / 2–output back-propagation network is given here. The operation of the network can be described by a simple recursive relationship between the outputs of each layer. At each node, a weighted sum of the outputs from all of the nodes of previous layers is acted on by a simple nonlinear function to provide the output for the node. A fixed, unity input is provided to each node to act as a threshold or bias. When the number of layers and nodes in each layer is chosen, the overall nonlinear function calculated by the network is determined by the values of the weights in the interconnections.



$$\mathbf{x}(k) = s\left[\mathbf{W}(k-1)\,\mathbf{x}(k-1)\right]$$

$$\mathbf{y} = \mathbf{x}(N) = f\left[\mathbf{x}(0)\right]$$

Exact Representation using Neural Networks

While investigating the limits of neural networks, Hecht-Nielsen [1987] was able to reinterpret a theorem by Kolmogorov related to the exact representation of a multi-input nonlinear function in terms of simple single-input functions. While back-propagation neural networks can be interpreted in terms of this theorem, the nonlinear function at each node is fixed by a limiting process which depends explicitly on the function to be represented. Most neural networks, on the other hand, have a common, simple function at the network node, and use the weighted interconnections to adjust the output of the network. There are many further difficulties with this theorem that limit its usefulness except as the most basic of existence proofs.

- Hecht-Nielsen's Interpretation of Kolmogorov's Theorem
  (Solution to Hilbert's 13th Problem)

  - Exact Representation

$$f(x) = \sum_{q=0}^{2n} g\left(\sum_{p=1}^{n} \lambda^{p-1} \psi(x_p + \varepsilon q) + q \right)$$

  - ψ can be obtained as a uniform limit of a sequence ψ_r of continuous nondecreasing piecewise linear functions
  - ψ is difficult to compute
  - ψ is different for each $f(x)$

Approximation Representation for Feedforward Networks

In most instances an approximate representation of a nonlinear function is all that is needed. Many researchers have been investigating the approximation capabilities of feedforward networks of the type described. These theorems are usually based on results from functional analysis and give sufficient conditions for the approximation of any continuous function to any desirable degree of accuracy. In general, they are not based on the Kolmogorov's Theorem for exact representation.

- Can feedforward networks approximate nonlinear functions?
- Recent theoretical results by
  - Funahashi,
  - Cybenko,
  - Hecht-Nielsen,
  - Stinchcombe and White,
  - and many others
- Based on results from functional analysis

Cybenko's Theorem

As an example of these recent theorems, the version due to Cybenko [1989] is outlined here. Defining a sigmoidal function in the most general form, Cybenko's Theorem shows that finite sums of sigmoidal functions, exactly the form of a single hidden layer feedforward neural network, are dense in the continuous functions on the unit hypercube. In simpler terms, any continuous function with inputs between 0 and 1 can be arbitrarily closely approximated by a neural network with one hidden layer. The other researchers have developed similar results. The differences are usually technical details related to how smooth the approximated function is and what type of functions are allowed in the nodes of the hidden layer.

While this is a very promising result, it is only an existence proof. The number of nodes necessary in the hidden layer is only specified as finite, and the weight vector is left unspecified.

Definition: We say that σ is *sigmoidal* if

$$\sigma(t) \to \begin{cases} 1 & \text{as } t \to +\infty \\ 0 & \text{as } t \to -\infty \end{cases}$$

Theorem: Let σ be any continuous sigmoidal function. Then finite sums of the form

$$g(x) = \sum_{j=1}^{N} \alpha_j \sigma(w_j^T x + \theta_j)$$

are dense in $C(I_n)$.

In other words, given $f \in C(I_n)$ and $\varepsilon > 0$, there is a sum, $g(x)$, of the above form, for which

$$|g(x) - f(x)| < \varepsilon$$

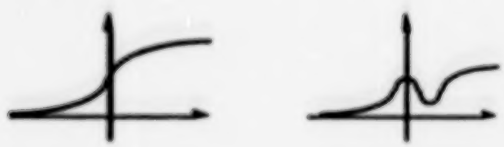
for all $x \in I_n$.

Example Functions

The sigmoidal functions, as defined for Cybenko's Theorem, can have many different shapes, including those shown here. The back-propagation algorithm requires a differentiable nonlinear function at each node. In most implementations the function is similar to the first one shown.

As mentioned previously, the other approximation theorems have different requirements on the nonlinearity allowed in the node. Two are shown here that fail to meet the requirements for Cybenko's Theorem, but are shown to be sufficient for function approximation by others.

- Satisfying Cybenko's Theorem



- Other Theorems

What's Ahead

Future work includes investigating the approximation capabilities of reasonable sized networks for multi-input functions of the type that will be found in a nonlinear control law. This will allow for a complete implementation of a nonlinear control law for the 737 aircraft using neural networks.

- Multidimensional Approximations
- Complete Nonlinear Inverse Dynamics implementation for 737
- Implement Neural Networks in NID control law
- Hope for constructive results regarding Neural Network function approximation

Stability Boundaries for Aircraft with Unstable Lateral-Directional Dynamics and Control Saturation**

Prakash C. Shrivastava* and Robert F. Stengel†
Princeton University, Princeton, New Jersey
**Published as J. Guidance, vol. 12, no. 1, Jan–Feb. 1989.

Aircraft that do not possess inherent (aerodynamic) stability must rely on closed-loop control systems for stable operation. Because there are limits on the deflections of an aircraft's control surfaces, the region of stable operation also is bounded. These boundaries are investigated for a lateral-directional example in which vertical fin size is inadequate to provide directional stability and where aileron and rudder deflections are subject to saturation. Fourth-order models are used in this study, with flight control logic based on minimum-control-energy linear-quadratic-regulatory theory. It is found that the stability boundaries can be described by *unstable* limit cycles surrounding stable equilibrium points. Variations in regions of stability with gain levels and command inputs are illustrated. Current results suggest guidelines for permissible limits on the open-loop instability of an aircraft's lateral-directional modes.

Introduction

THE performance and flying qualities of future aircraft will be greatly affected by advanced flight control systems, which can augment stability and command response over the entire flight envelope. Aerodynamic stability need not be built into an aircraft if control effectiveness and power are sufficient to maintain stability by feedback. Unstable aircraft depend on their flight control systems to achieve stability, but constraints on the magnitudes of control deflections limit the region of stable operation. The size and shape of the region of stability depend on the degree of instability, the maximum control forces and moments available, and the feedback control logic.

In Ref. 1, an analysis and design of a stability augmentation system (SAS) for an aircraft with longitudinal static instability and elevator saturation are presented. Analysis of control saturation in command response of statically unstable aircraft is presented in Ref. 2. In these investigations, the stability and saturation boundaries were found to be parallel if a minimum-control-energy (MCE) linear-quadratic (LQ) control law was used. However, the investigation was limited to the case of a single saturating control with a saddle-point phase-plane singularity. The stability boundaries for single-input systems with other types of singularities, e.g., unstable nodes and unstable foci, are presented in Refs. 3 and 4. These analyses were performed using normal-mode coordinates, which simplified determination of the stability boundaries. The stability boundaries were shown to depend on the type of singularities and the magnitude of the feedback gains. For unstable nodes and foci, the stability boundaries were shown to be unstable limit cycles.

This paper examines the effects of control saturation on unstable lateral-directional dynamics. Linear dynamic models of lateral-directional motion are reviewed. Stability boundaries are compared for two cases of directional instability. Since high-gain controllers often are used to increase the system performance, effects of high gains on the regions of stability are examined. The control saturation analysis is extended to determine the changes in stability boundaries with commands.

Presented as Paper 85-1948 at the AIAA Guidance, Navigation, and Control Conference, Snowmass, CO, Aug. 19-21, 1985; received Sept. 9, 1985; revision received Oct. 9, 1987. Copyright © American Institute of Aeronautics and Astronautics, Inc., 1987. All rights reserved.

*Doctoral Candidate, Department of Mechanical and Aerospace Engineering; currently, Member of the Advanced Engineering Staff, General Motors Corporation.

†Professor, Department of Mechanical and Aerospace Engineering. Associate Fellow AIAA.

Fundamental Analysis of Control Saturation

Control displacement saturation effects are considered using a linear perturbation model of the form

$$\dot{x} = Fx + Gu \tag{1}$$

The state vector x contains the yaw rate r, sideslip angle β, roll rate p, and roll angle ϕ; the control vector u contains rudder and aileron positions δr, δa. The matrices F and G contain dimensional stability derivatives[3] as defined below. In the examples considered here, F has eigenvalues with positive real parts, i.e., the system is open-loop unstable, and the control variable u is constrained to lie within the limits $\pm u_m$.

$$F = \begin{bmatrix} N_r & N_\beta & N_p & 0 \\ -1 & Y_\beta/V_0 & 0 & g/V_0 \\ L_r & L_\beta & L_p & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{2}$$

$$G = \begin{bmatrix} N_{\delta r} & N_{\delta a} \\ 0 & 0 \\ L_{\delta r} & L_{\delta a} \\ 0 & 0 \end{bmatrix} \tag{3}$$

Closed-Loop Control Effects

LQ control theory is used to specify a stability augmentation system.[5] The control law obtained is of the form

$$u = -Cx = \begin{bmatrix} \delta_r \\ \delta_a \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \end{bmatrix} \begin{bmatrix} r \\ \beta \\ p \\ \phi \end{bmatrix} \tag{4}$$

This LQ closed-loop regulator is guaranteed to remain stable as long as all the controls remain unsaturated; however, no such guarantee can be made if one or both of the controls are subject to saturation. With no command inputs, trajectories beginning at certain locations in the state space will converge to the origin, which is a stable equilibrium point of the unsaturated system, whereas other trajectories will diverge. Stability regions can include regions of saturated control, whereas regions of instability can include regions of unsaturated control.

The boundaries between saturated and unsaturated regions are found directly from Eq. (4). Each fixed (saturated) level of scalar control defines a hyperplane separating saturated and unsaturated regions in the state space; hence, there are nine distinctly different regions corresponding to no-saturation or to positive- or negative-saturation of one or both controls in this two-control example. Within these regions, there are up to nine different equilibrium points and four sets of eigenvalues and eigenvectors. For zero command input, asymptotic stability is defined as the tendency for trajectories to approach the single equilibrium point at the origin; therefore, stable trajectories may pass through regions with positive eigenvalues, and unstable trajectories may transit regions with negative eigenvalues. Whether or not a trajectory is convergent depends on the location of equilibrium points, as well as the signs of the eigenvalues. The same is true for nonzero command inputs, except that the stable equilibrium point is no longer at the origin.

MCE Control Law

The LQ control law approaches the MCE solution (in the limit) as the diagonal elements of the control weighting matrix approach infinity. High control weighting leads to low feedback gains and so saturation tends to occur at high levels of feedback state magnitude. Because the cost on control usage is very high, the stable poles are not shifted by the feedback control, but the unstable poles are reflected to their stable left-half s-plane images.[6]

Normal-Mode Coordinates

Analysis of control saturation effects is facilitated by transforming the original state vector into the normal-mode coordinates of the open-loop system. The transformation to the normal-mode coordinate system is obtained by

$$x = Mq \tag{5}$$

where M is the modal matrix and q the normal-mode vector associated with the diagonal system $M^{-1}x$. The corresponding dynamic equation is

$$\dot{q} = \Lambda q + \Xi u \tag{6}$$

where Λ is a diagonal matrix of eigenvalues diag $(\lambda_1, \lambda_2, \ldots \lambda_n)$, and $\Xi = M^{-1}G$.

This mathematical model can be written as

$$\begin{bmatrix} \dot{q}_u \\ \dot{q}_s \end{bmatrix} = \begin{bmatrix} F_u & 0 \\ 0 & F_s \end{bmatrix} \begin{bmatrix} q_u \\ q_s \end{bmatrix} + \begin{bmatrix} G_u \\ G_s \end{bmatrix} u \tag{7}$$

where q_u and q_s represent unstable and stable modal states, the matrices F_u and F_s the unstable and stable dynamics, and G_u and G_s the corresponding control matrix components.

The MCE control law in the normal-mode space is of the form

$$u = [-C_u \; 0] \begin{bmatrix} q_u \\ q_s \end{bmatrix} \tag{8}$$

Substitution of this control law into Eq. (7) yields the lower-block-triangular closed-loop system

$$\begin{bmatrix} \dot{q}_u \\ \dot{q}_s \end{bmatrix} = \begin{bmatrix} F_u - G_u C_u & 0 \\ -G_s G_u & F_s \end{bmatrix} \begin{bmatrix} q_u \\ q_s \end{bmatrix} \tag{9}$$

which can also be written as

$$\dot{q}_u = (F_u - G_u C_u) q_u \tag{10}$$

$$\dot{q}_s = F_s q_s - G_s C_u q_u \tag{11}$$

Eigenvectors associated with all the modes are changed. The equations of the unstable modes remain uninfluenced by the stable modes, but the stable mode equation has a driving term that consists of the states of the unstable mode and the feedback gain. In the steady state, the origin represents the stable equilibrium point for this system. Since there is no contribution from the stable modes to the trajectories of the unstable modes, the saturation and stability boundaries of the entire system are confined to the plane of unstable mode coordinates. Taken two components at a time, projections of trajectories and intersections of saturation/stability boundaries can be plotted in conventional phase-plane fashion. Single-unstable-root cases lead to open stability boundaries, whereas a pair of real or complex unstable roots causes closed boundaries. Once the boundaries are determined in the normal-mode space, they can be transformed back to the original state space using the inverse transformation.

Determination of Phase-Plane Stability Boundaries

Consider the normal-mode equations with positive saturated control $+u_m$, given by

$$\dot{q}_1(t) = \lambda_1 q_1(t) + b_1 u_m \tag{12}$$

$$\dot{q}_2(t) = \lambda_2 q_2(t) + b_2 u_m \tag{13}$$

The above system has a saddle-point singularity when only one of the eigenvalues is positive, and it has an unstable node singularity when both the eigenvalues are positive. The equilibrium values can be obtained by setting the derivatives $\dot{q}_1$ and $\dot{q}_2$ to zero. Then, q_1^* and q_2^* are

$$q_1^* = -b_1 u_m / \lambda_1 \tag{14}$$

$$q_2^* = -b_2 u_m / \lambda_2 \tag{15}$$

Because the LQ control law can be written as

$$u(t) = -c_{11} q_1(t) - c_{12} q_2(t) \tag{16}$$

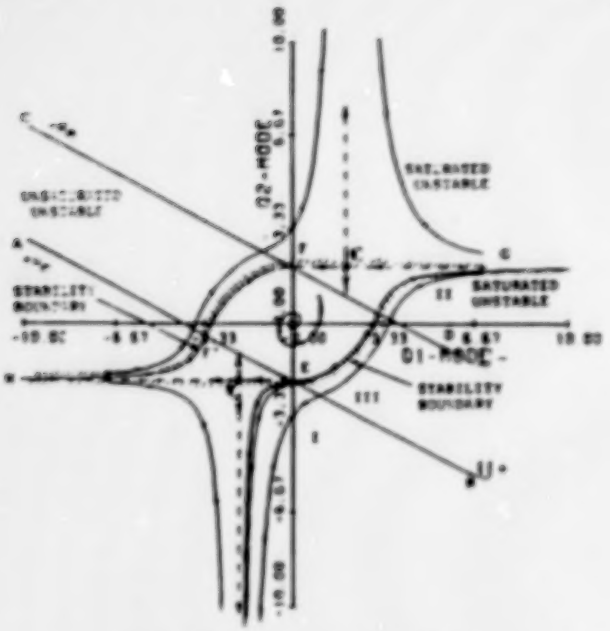
the normal-mode saturation boundaries are given by the above equation, substituting $u = \pm u_m$. Since Eqs. (12) and (13) are decoupled, each can be solved independently, yielding

$$q_1(t) = [q_{10} + (b_1 u_m)/\lambda_1 e^{\lambda_1 (t)}] + (b_1 u_m)/(-\lambda_1) \tag{17}$$

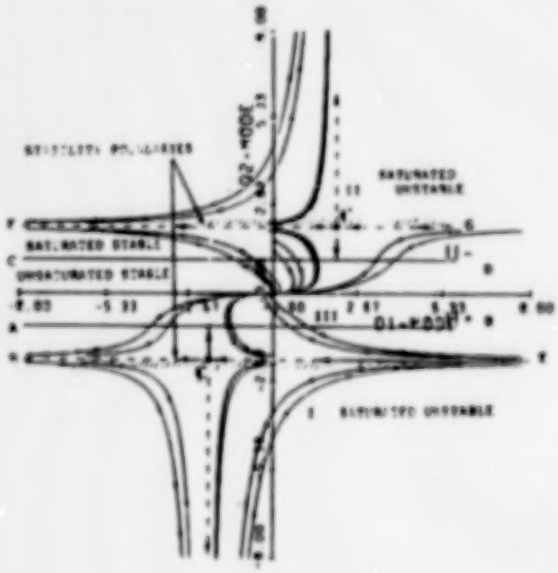
$$q_2(t) = [q_{20} + (b_2 u_m)/\lambda_2 e^{\lambda_2 (t)}] + (b_2 u_m)/(-\lambda_2) \tag{18}$$

Saddle-Point Case

The saddle-point case is obtained when $\lambda_1 < 0$ and $\lambda_2 > 0$. For those initial conditions that make the signs of the coefficient of the exponent corresponding to the unstable mode λ_2 [Eq. (18)] and the constant term differ, the trajectories converge toward the saturation boundary; the trajectories diverge when these terms have the same sign. The system remains at its equilibrum value for those initial conditions where the coefficient of the exponent in Eq. (18) vanishes. This requirement is met by the axes of the unstable modes passing through the saturated equilibrium points, which are always located in the saturated regions[4]; hence, they represent parts of the stability



a) Feedback of stable and unstable modes



b) Feedback of unstable mode only

Fig. 1 Stability boundaries for saddle-point case (normal-mode coordinates).

boundaries in the saturated regions. When both the modal coordinates are fed back, the saturation boundaries intersect the axes of the normal modes.

The portions of the stability boundaries through the unsaturated regions are determined by the dynamics of the closed-loop system using the points of intersections of the saturation boundaries and the axes of the unstable mode, as illustrated in Fig. 1a. When only the unstable mode is fed back, the saturation boundaries are parallel to the unstable-mode axes emanating from the saturated equilibrium points. These axes represent the complete stability boundaries in the normal-mode coordinates, as shown in Fig. 1b.

Unstable Node Case

The unstable node case is obtained when λ_1 and λ_2 are positive. For this system to be stabilized, both modes must be fed back; hence, the saturation boundaries always would be inclined to the normal-mode axes, and stability boundaries

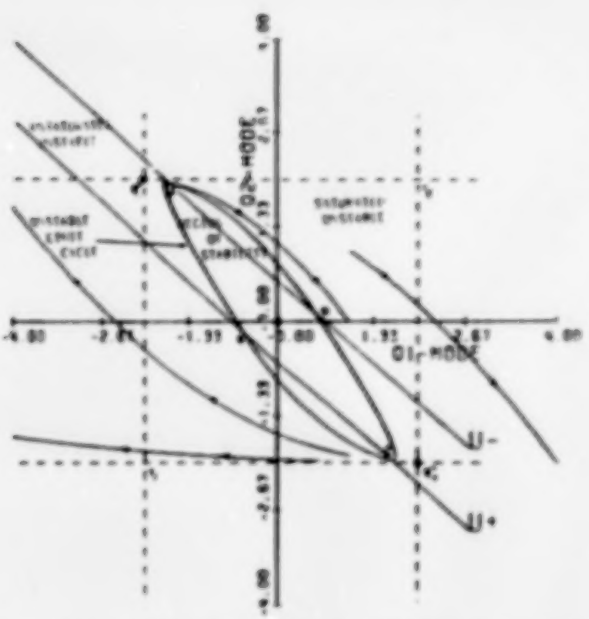


Fig. 2 Stability boundaries for unstable node case (normal-mode coordinates).

similar to Fig. 1b are ruled out. Since the equilibrium points due to control saturation are always located outside the saturated region,[4] boundaries similar to Fig. 1a are not possible. With $\lambda_i > 0$, the trajectories converge when the signs of each coefficient of the exponents and the constant terms in Eqs. (17) and (18) are opposite. In this case, only parts of the q_1 and q_2 axes within the equilibrium values contain stable regions (Fig. 2).

The rectangle formed by the equilibrium axes establishes the outside limit for the stability region. There is a stable singular point within the linear region at the origin, and neighboring trajectories can be forced to the origin. Therefore, one can anticipate existence of an unstable limit cycle that encloses a stable equilibrium point. Any small perturbation of the motion from this limit cycle would result in a departure to either the unstable region or the origin. A general closed-form analytical expression for this limit cycle cannot be determined; hence, an algebraic test to determine whether an initial condition lies within or outside the limit cycle is difficult to find. However, the unstable limit cycle can be obtained numerically by integrating the system of equations backward in time, with initial conditions in the vicinity of the origin.

Effects of Command Inputs

Command augmentation systems (CAS) are designed to achieve desired equilibrium values and suitable system response to external commands. The CAS commands the state to nonzero equilibrium value, and it may modify transient response by closed-loop control.[7] In this study the desired output is represented by

$$y(t) = H_x x(t) \tag{19}$$

and so state equilibrium implies command equilibrium as well:

$$y^* = H_x x^* \tag{20}$$

Following Ref. 7, the nonsingular command augmentation control law can be written as

$$u(t) = C_f y^* + C_b x(t) \tag{21}$$

with

$$C_f = (-H_x F^{-1} G)^{-1} + C(-F^{-1} G)(-H_x F^{-1} G)^{-1} \tag{22}$$

$$C_\beta = -C \tag{23}$$

The system's closed-loop response to the command is described by

$$\dot{x}(t) = (F + GC_\beta)x(t) + GC_\gamma y^* \tag{24}$$

The eigenvalues and eigenvectors of Eq. (24) are associated only with $(F + GC_\beta)$, and so the choice of the command vector has no effect on these parameters. From Eq. (21) it is clear that the saturation boundaries vary with changes in the command vector; therefore, the stability boundaries also vary. Nonzero command vectors yield saturation boundaries that are asymmetric with respect to the origin. The feedback gain C_β affects only the closed-loop eigenvalues; it has no effect on the value of u^* for a desired command. Control saturation limits $\pm u_m$ determine the system's ability to achieve the command vectors in the steady state. Desired outputs (i.e., constant commands) that cause the control to violate the limits

$$-u_m \le u^* \le +u_m$$

cannot be attained. Consequently, the minimum and maximum values of the equilibrium states for the above saturation limits are given by

$$x^*_{max/min} = \pm F^{-1}Gu_m \tag{25}$$

and the maximum/minimum values of command vectors that can be commanded and achieved are given by

$$y^*_{max/min} = H_x x^*_{max/min} \tag{26}$$

In state-space coordinates, x^* lies on a hyperplane determined by Eq. (25); therefore, command vectors produce equilibrium states that lie on this hyperplane. At control saturation limits $\pm u_m$, the saturated and unsaturated equilibrium points sought by the command vector coincide, and they lie on the saturation boundary. Variations in the command vector change the location of the desired equilibrium point and saturation boundaries. The locations of equilibrium points with all controls saturated do not change, but the locations of equilibrium points with one of the controls saturated change with the commands. Because of this change in the single-control-saturated equilibrium points relative to fixed all-control-saturated equilibrium points, the shapes and sizes of the stability boundaries change with the command vector.

The maximum and minimum values of attainable state equilibrium depend upon the control saturation limits and the open-loop dynamics, and control saturation limits set bounds on the achievable command vector y^*. Although achievable command vectors are independent of the feedback gain C_β, transient response can impact the region of achievable output vectors. If, for example, the transient response to an "achievable" command momentarily forces the state into a divergent region, then the trajectories can diverge and that command is not truly achievable.

Effects of Vertical-Fin-Size Reduction on Lateral-Directional Dynamics

In conventional aircraft configurations, the lateral-directional dynamics normally are characterized by the Dutch roll, roll subsidence, and spiral modes; however, reducing the vertical fin size alters the aircraft's response modes.[9,9] With sufficient reduction in fin size, the vertical tail contribution to the static directional stability derivative is reduced, making the nondimensional directional stability derivative C_{n_β} negative. The yaw damping derivative C_{n_r}, control power derivative $C_{n_{\delta_a}}$, and dihedral derivative C_{l_β} are reduced in magnitude, whereas the roll derivatives and roll-to-yaw coupling derivatives remain relatively unaffected. The effects on dimensional stability derivatives (e.g., N_β) are proportional to the changes in the nondimensional derivatives (e.g., C_{n_β}). The major impact of fin-size reduction is felt on N_β, which becomes negative, making the aircraft directionally unstable. In Ref. 10 it is shown that there always would be two unstable roots.

Numerical Models for Analysis

A high-performance fighter aircraft with stable lateral-directional dynamics was chosen as a baseline model for this study. Nondimensional derivatives for the unstable lateral-directional model were computed for several fin configurations, assuming a sea level, Mach 0.4 flight condition. The dimensional stability derivatives for three fin configurations are presented in Table 1. Case 1 represents the stable baseline model, and cases 2 and 3 represent the unstable models obtained by the fin-size reduction. The numerical values assume that velocity and angles are measured in feet per second and radians, respectively. The rudder and aileron limits are fixed at ± 20 deg for all numerical calculations. The eigenvalues of the open-loop dynamics for the fourth-order models are given in Table 1.

Table 1 Dimensional derivatives eigenvalues for three aircraft configuration

Derivative	Case 1	Case 2	Case 3
$N_{\dot{\beta}}$	−0.379	−0.209	−0.177
N_β	3.740	−0.865	−2.929
N_p	−0.104	−0.072	−0.070
Y_β/V_0	−0.258	−0.207	−0.195
L_r	0.736	0.506	0.474
L_β	−11.98	−5.87	−4.98
L_p	−2.51	−2.49	−2.48
N_{δ_r}	−6.21	−3.53	−2.96
N_{δ_a}	0.287	0.287	0.289
L_{δ_r}	5.54	1.96	1.39
L_{δ_a}	20.12	20.12	20.12
Eigenvalues	$-0.207 + j2.074$	−0.857	−1.396
	$-0.207 - j2.074$	0.459	1.143
	−2.72	−2.642	−2.647
	−0.012	0.133	0.0363

Table 2 Minimum-control-energy gains for normal-mode and state coordinates

	Normal-mode coordinates		State-space coordinates			
	q_1	q_2	r	β	p	ϕ
Case 2						
δr	−0.39	−0.15	−0.306	0.246	0.015	0.026
δa	−0.04	0.02	−0.057	0.036	0.005	0.009
Case 3						
δr	−1.33	−0.06	−0.737	0.5	0.032	0.066
δa	−0.24	0.06	−0.141	0.183	0.009	0.019

Effects of Control Saturation on Closed-Loop Lateral-Directional Stability

Linear-quadratic gains in normal-mode coordinates and physical coordinates were computed for the unstable lateral-directional dynamics of case 2 and case 3 configurations. The state weighting matrix was an identity matrix, and the control weightings for the ailerons and rudder were fixed at $R = \text{diag}(500{,}000;500{,}000)$, approximating the MCE case. The MCE gains are given in Table 2. Gains for the rudder controls are higher than gains for aileron control, because the low damping and directional stability caused by small fin size are highly augmented by the rudder control law. Variations in the roll-rate and bank-angle feedback gains to the rudder are small, because the effect of fin-size reduction on the roll mode is small. Eigenvalues for the closed-loop system and saturated rudder or ailerons are given in Table 3.

Figure 3a presents the saturation and stability boundaries in open-loop normal-mode coordinates (q_1,q_2) for case 2. The rudder and aileron saturation boundaries are straight lines, and the stability boundary is a closed curve, an unstable limit cycle. The region enclosed by the limit cycle is the region of stability, and all initial conditions in it converge to the origin. Initial conditions that fall outside this region result in divergence. Note that this limit cycle lies within the saturation boundaries of the aileron; for the given aileron saturation limits, aileron control does not influence the shape or size of the region of stability. For the region of stability to be influenced, either increased aileron feedback gain or reduced aileron saturation control limit is necessary.

Table 3 Eigenvalues with minimum-control-energy gains

	Closed-loop	Rudder saturated	Ailerons saturated
Case 2	−0.459	0.41	0.26
	−0.133	0.07	0.93
	−0.857	−0.85	−1.36
	−2.566	−2.64	−2.64
Case 3	−1.143	1.10	4.56
	−0.0363	−0.11	$-1.2+j0.14$
	−1.396	−1.90	$-1.12-j0.14$
	−2.647	−3.34	−2.64

For the MCE control laws, saturation boundaries in the four-dimensional state space are hyperplanes parallel to the axes of the stable modal coordinates q_3 and q_4. Since the stable modes are unaffected by the MCE control laws, the stability region can be described in the $(q_1\text{-}q_2)$ plane; hence, it is described by a hypercylinder in the four-dimensional normal-mode space. This hypercyclinder does not intersect the axes of the stable modes, and it transforms to an equivalent hypercylinder in the original state space, as shown below.

The linear transformation given by Eq. (5) can be rewritten as

$$\begin{bmatrix} r \\ \beta \\ p \\ \phi \end{bmatrix} = \begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} \tag{27}$$

In Eq. (27), q_1 and q_2 are the points on the limit cycle. Given q_1, q_2, β, and ϕ, the remaining normal-mode coordinates q_3 and q_4 are readily obtained by

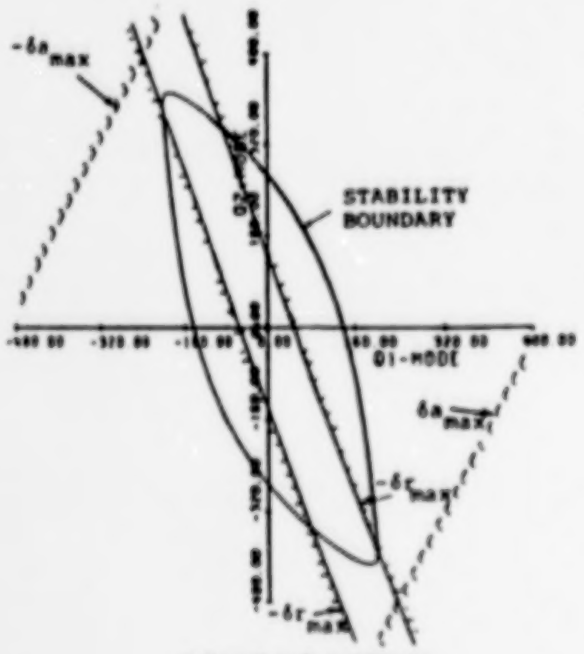
$$\begin{bmatrix} q_3 \\ q_4 \end{bmatrix} = T_4^{-1} \left\{ \begin{bmatrix} \beta \\ \phi \end{bmatrix} - T_3 \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \right\} \tag{28}$$

Substituting Eq. (28) in Eq. (27) then yields the region of stability in the r-p plane, which also is a closed region. The stability boundaries for nonzero sideslip angles are shown in Fig. 3b. Sideslip angle is seen to have a small effect on the stability region, as is true for roll angle (not shown).[10]

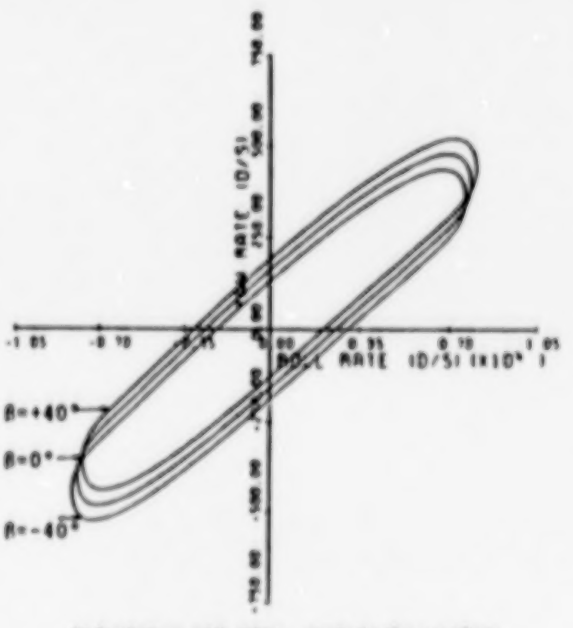
Effects of Gain Variations

High-feedback-gain controllers normally are used to reduce the effects of parameter variations, and high-forward-loop gains quicken vehicle response. As a result of high-feedback gain, the closed-loop poles move farther into the left-half s plane; control saturation occurs for relatively small values of state variables, and the region of unsaturated operation is reduced.

To illustrate the effect of high gains on stability boundaries, MCE gains for the case 2 configuration were increased arbitrarily by a factor of five. (The new feedback gains do not



a) Normal-mode coordinates



b) Variations with sideslip angle in the r-p plane

Fig. 3 Stability boundaries for case 2 with minimum-control-energy gains.

minimize the original cost function.) Eigenvalues for the closed-loop system and saturated rudder or aileron are given in Table 4. The eigenvalues for MCE and high-gain cases in Tables 3 and 4 show that the high-feedback-gain system is relatively more stable even with saturated rudder or ailerons. Intuitively, the size of the stability region should also increase with increased system stability. The validity of this intuition is illustrated in this example. The saturation and stability boundaries in normal-mode and physical coordinates are shown in Fig. 4. By comparison to Fig. 3a, the size of the stability region is seen to be increased; however, the stability boundary in Fig. 4 passes through the saturated-aileron region, and so the stability region depends on the aileron saturation limits.

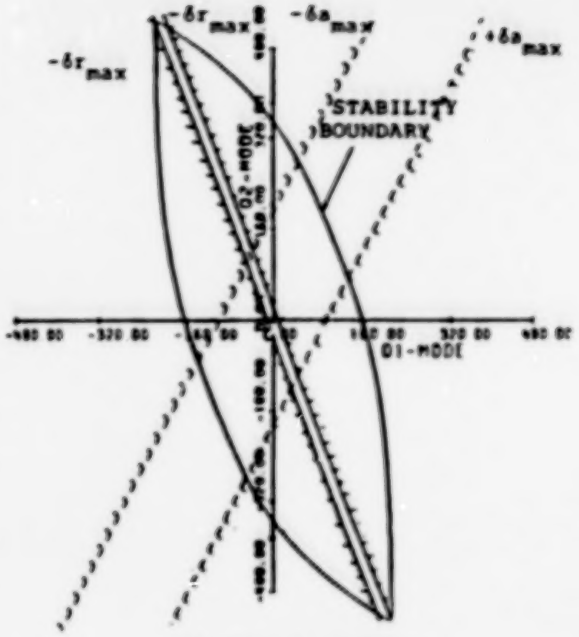
Consider the case when elements in the control weighting matrices for the rudder and aileron are not the same. High (low) control weighting on rudder and low (high) control weighting on aileron result in low (high) feedback gains for the rudder and high (low) feedback gains for the aileron, respectively. Although not MCE, high values of control weightings ensure that there is negligible alteration of stable modes. In such cases, the normal-mode stability regions also are hypercylinders parallel to the axes of the stable modes. Thus, the regions of stability in the r-p planes still can be determined using Eqs. (27) and (28).

The effects of low rudder and high aileron weightings (and vice versa) on the stability boundaries are examined, beginning with rudder weighting of 500,000, aileron weighting of 5000, and $Q = I$ (Table 5). Even though the rudder gains are small, saturating the rudder produces a right-half-plane eigenvalue; however, rudder saturation is not likely to be a problem, as shown by the saturation and stability boundaries of Fig. 5. Saturating the ailerons also produces a right-half-plane eigenvalue, indicating that the nominal, two-control system makes use of the yawing moments for both controls. The stability boundaries remain unaffected by the rudder saturation boundary, and the size of this stability region is smaller than that shown in Fig. 3.

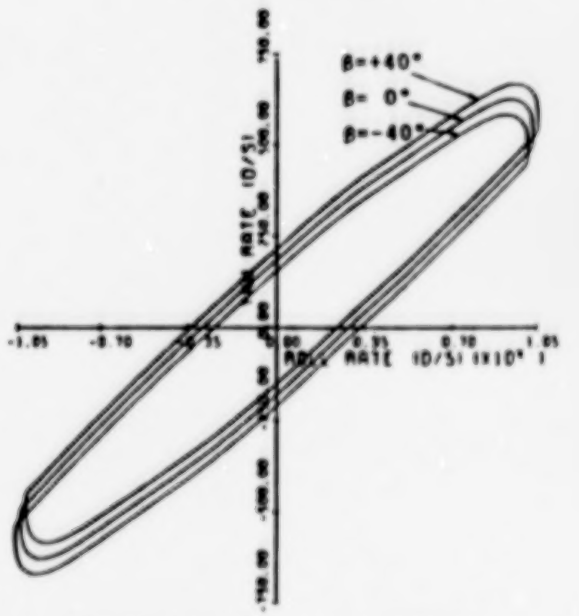
The gains and eigenvalues for rudder weighting of 5000 and aileron weighting of 500,000 are given in Table 6, and the

Table 4 Eigenvalues with five times minimum-control-energy gains

	Closed-loop	Rudder saturated	Ailerons saturated
Case 2	−5.164	0.325	0.249
	−0.174	−0.333	−0.882
	−2.642	−0.818	−2.645
	−0.174	−2.682	−8.443
Case 3	−0.221	0.997	0.306
	−1.396	−0.803	−1.493
	−2.646	−2.778	−2.668
	−10.40	−1.283	−6.778



a) Normal-mode coordinates



b) Variations with sideslip angle in the r-p plane

Fig. 4 Stability boundaries for case 2 with five times minimum-control-energy gains.

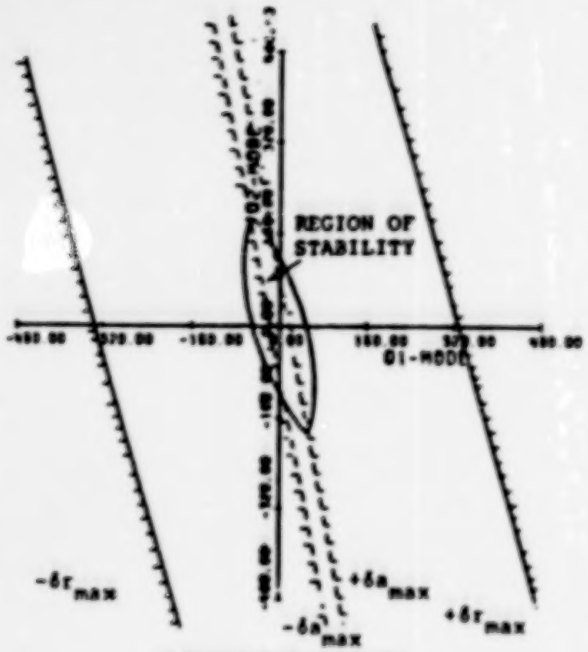
Table 5 Control gains and eigenvalues for case 2 with high rudder weighting

Gains

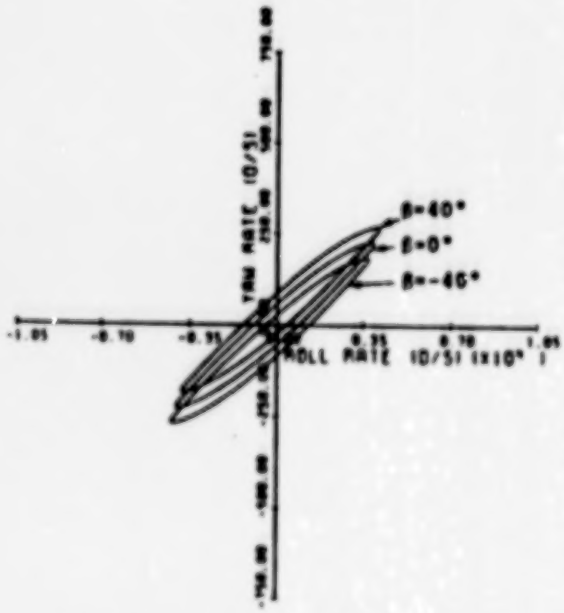
	Normal-mode coordinates			State vector coordinates		
	q_1	q_2	r	β	p	ϕ
δr	−0.06	−0.02	−0.05	0.039	0.003	0.005
δa	−1.23	−0.26	−1.11	0.836	0.065	0.12

Eigenvalues

Closed-loop	Rudder saturated	Ailerons saturated
−.135	0.30	0.33
−4.6	−2.97	−0.04
−2.64	−1.86	−0.85
−0.86	−0.91	−2.64

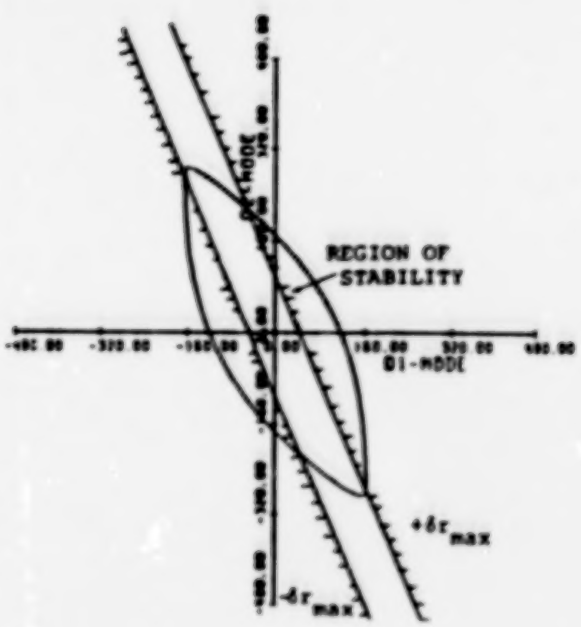


a) Normal-mode coordinates

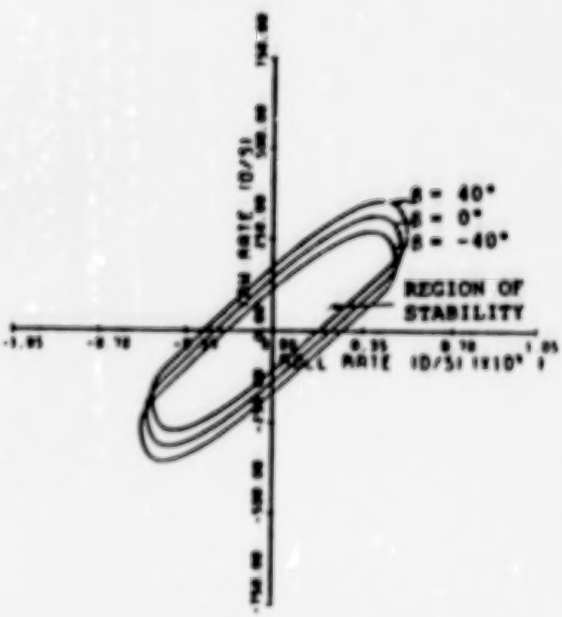


b) Variations with sideslip angle in the r-p plane

Fig. 5 Stability boundaries with high rudder weighting for case 2; R = diag(500,000;5000).



a) $y^* = (0.0,0.0)$



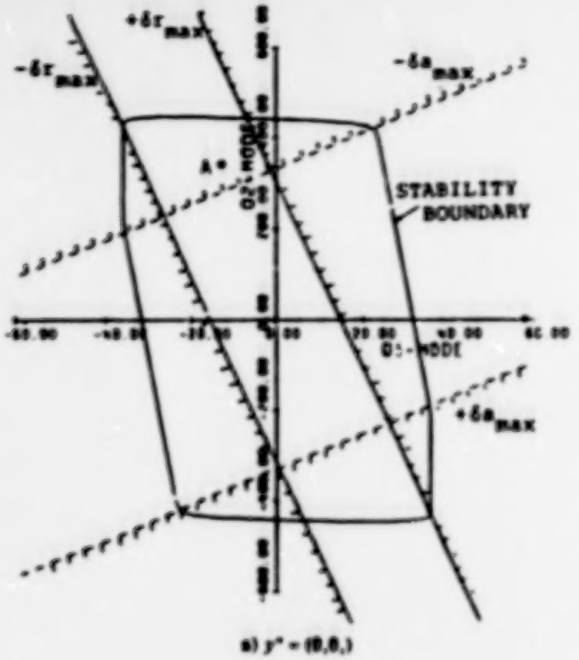
b) Variations with sideslip angle in the r-p plane.

Fig. 6 Stability boundaries with high aileron weighting for case 2; R = diag(5000;500,000).

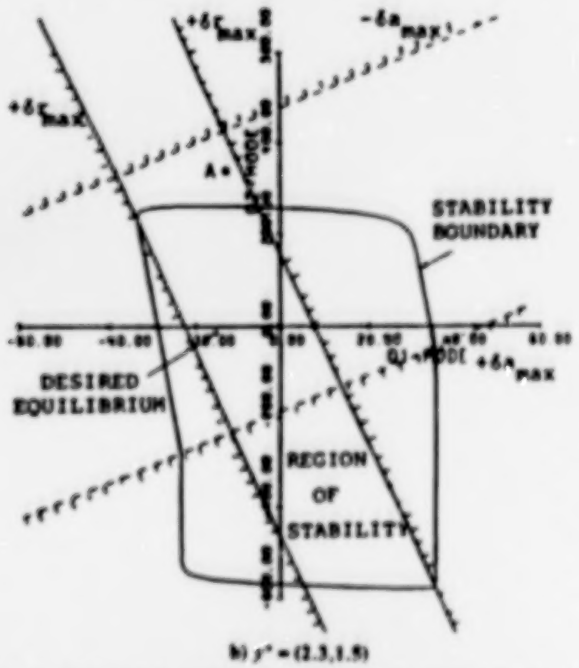
Table 6 Control gains and eigenvalues for case 2 with high aileron weighting

	Gains					
	Normal-mode coordinates			State vector coordinates		
	η_1	η_2	r	β	p	ϕ
δr	−0.44	−0.19	−0.33	0.27	0.013	0.025
δa	−0.0002	0.0004	−0.0005	0.0003	0.00005	0.00001

Eigenvalues		
Closed-loop	Rudder saturated	Aileron saturated
−0.14	0.46	0.26
−0.46	0.13	−1.53
−2.64	−2.64	−0.91
−0.86	−0.86	−2.64



a) $y^* = (0,0,)$



b) $y^* = (2.3,1.5)$

Fig. 7 Variation in stability boundaries with commands; case 3, minimum-control-energy gains (normal-mode coordinates).

saturation and stability boundaries are shown in Fig. 6. The region of stability is larger in size than for the high-rudder-weighting case because high-feedback gains to the rudder provide better directional stability augmentation than feedback to the ailerons, and the stability boundary is invariant with respect to aileron saturation limits. For both cases of unequal control weightings, the variation in the region of stability with sideslip angle is small (Figs. 5b and 6b). Because feedback gains from bank angle to the rudder and ailerons are small, the effect of bank angle on the region of stability is small.

Comparison of the stability boundaries in Figs. 3–7 shows that the largest regions of stability are obtained when the control weightings on the rudder and aileron are equal and that the size of the stability region increases when the feedback gain magnitudes increase. The smallest region of stability is obtained for the high-rudder-weighting case. Nevertheless, although high-feedback gains are desirable for achieving a large stability region, these gains also must satisfy criteria for normal operation in the unsaturated region, which may limit the acceptable gain magnitude. Except for the case of high rudder weighting, the variations in stability regions with sideslip angles are small. Sideslip angle dependency could be ignored if rolling and yawing rates could be restricted to the innermost intersection of all stability regions.

Effects of Command Inputs

Commanded inputs have an effect on system stability because they offset the desired equilibrium point from the origin, and stability boundaries become asymmetrical. The desired equilibrium point may lie close to a stability boundary, and so small disturbances can make the system unstable. This is demonstrated for case 3 using a nonsingular "task-tailored" command vector, in which yaw rate and sideslip angle are commanded independently. (A more conventional command vector would be roll rather than yaw rate; this is a singular-command variable in the sense of Ref. 7.) The output matrix is then $H_y = \text{diag}(1,1)$ and, with the feedback-gain matrix in Table 3, the forward-gain matrix is

$$C_f = \begin{bmatrix} 0.378 & 5.762 \\ 0.278 & 5.412 \end{bmatrix} \tag{29}$$

The desired equilibrium states must lie within the saturation limits which, themselves, shift with the value of the command.

Two sets of command vectors are considered. In the first instance (Fig. 7a) the commands are zero. The saturation boundaries are symmetric about the origin, and the stability boundary generated by backward integration is antisymmetric about the normal axes. Case 3 is more unstable than case 2, and its stability region is, therefore, smaller. Its stability boundary passes through regions in which both rudder and ailerons are saturated.

When the command is changed to $y^* = (2.3,1.5)$, the stability boundary becomes asymmetrical, and no portion of it passes through the region of negative aileron satuation (Fig. 7b). Furthermore, portions of the space that were stable with zero command are no longer stable regions. For example, if a disturbance forces the instantaneous (q_1,q_2) point to be point A as the pilot commands $y^* = (2.3,1.5)$, the system is unstable because point A does not lie within the stable region of that command. It can be concluded that the allowable perturbation space is smaller than the zero-command stability boundary when nonzero commands are issued. Figure 7b also indicates that the allowable command space is smaller than the unsaturated region of the system.

Conclusions

General conditions for stability boundaries of linear systems containing control saturation have been presented. For linear control laws, saturation boundaries always appear as hyperplanes in the normal-mode space. The shape of the stability boundaries depends on the type of open-loop instability. Fourth-order models of lateral-directional aircraft motion exhibit two unstable roots that result from the lack of directional stability. Stability regions for aircraft with directional instability are found to be hypercylinders in the normal-mode space when only unstable modes are fed back; the stable regions can be increased in size by increasing the magnitudes of feedback gain. These boundaries transform to hypercylinders in the original state space. For small directional instabilities, variations in stability regions with sideslip and roll angles are small, and the innermost stability region in the roll rate-yaw rate phase plane can represent the overall region of stability. With increased directional instability, the stability region is reduced, and variations in the regions of stability with sideslip angle become significant. Stability regions are command-dependent, and a possibility of departure exists if

initial conditions at the time of new command initiation lie outside the stability region.

Acknowledgments

This research was sponsored by NASA Langley Research Center, Hampton, Virginia, under Grant NGL 31-001-252.

References

[1]Hanson, G.D. and Stengel, R.F., "Effects of Displacement and Rate Saturation on the Control of Statically Unstable Aircraft," *Journal of Guidance, Control, and Dynamics*, Vol. 7, March-April 1984, pp. 197-205.

[2]Hanson, G.D. and Stengel, R.F., "Effects of Control Saturation on the Command Response of Statically Unstable Aircraft," AIAA Paper 83-0065, Jan. 1983.

[3]Shrivastava, P.C. and Stengel, R.F., "Stability Boundaries for Systems with Control Constraints," *Proceedings of the 18th Annual Conference on Information Sciences and Systems*, Princeton Univ., Princeton, NJ, March 1984.

[4]Shrivastava, P.C. and Stengel, R.F., "Stability Boundaries for Closed-Loop Systems with Control Constraints," *Proceedings of the 23rd IEEE Conference on Decision and Control*, IEEE, New York, Dec. 1984, pp. 1326-1329.

[5]Seckel, E., *Stability and Control of Airplanes and Helicopters*, Academic Press, New York, 1964.

[6]Stengel, R.F., *Stochastic Optimal Control: Theory and Application*, Wiley, New York, 1986.

[7]Stengel, R.F., "Equilibrium Response of Flight Control Systems," *Automatica*, Vol. 18, March 1982, pp. 343-348.

[8]Roskam, J., *Methods for Estimating Stability and Control Derivatives of Conventional Subsonic Airplanes*, Univ. of Kansas, Lawrence, KS, 1972.

[9]Etkin, B., *Dynamics of Flight: Stability and Control*, Wiley, New York, 1959.

[10]Shrivastava, P.C., "Stability Regions of Relaxed Static Stability Aircraft Under Control Constraints," Ph.D. Thesis, Mechanical and Aerospace Engineering Dept., Princeton Univ., Princeton, NJ, Rept. 1747-T, Oct. 1986.

Restructurable Control Using Proportional-Integral Implicit Model Following*

Chien Y. Huang* and Robert F. Stengel†
Princeton University, Princeton, New Jersey 08544
**Published as *J. Guidance*, vol. 13, no. 1, March–April 1990.

Studies of a proportional-integral implicit model-following control law are presented. The research focuses on the ability of the control law to recover the performance of a system with failed actuators or structural damage to its prefailure level. Properties of the implicit model-following strategy are examined, and conditions for control reconfiguration are stated. The control law is applied to the lateral-directional model of a fighter aircraft, and control restructuring is shown for changes in control and system matrices. It is concluded that the implicit model-following scheme is a good candidate for control reconfiguration.

Introduction

FAULT-TOLERANT control systems typically have one or more of three attributes: robustness, reconfigurability, and restructurability. Robustness implies insensitivity to variations in the parameters or structure of the system. If physical parameters change or the nature of system dynamics is altered, the robust system continues to operate as originally intended, providing response to inputs and disturbances that is close to nominal performance. Reconfigurability implies that some of the control system's parameters can be purposely modified to account for uncontrollable changes in the system, such as failed sensors, actuators, or damaged structural elements. Restructurability subsumes reconfigurability, implying that not only parameters but the structure of the system itself can be changed to accommodate uncontrollable changes. Whereas the robust system obtains fault tolerance with fixed parameters and structure, reconfigurable and restructurable systems are adaptive, providing some degree of fault detection and identification as well as the ability to alter the system. The latter should achieve satisfactory performance over a wider range of conditions than is obtained by robustness alone; otherwise the added complexity is not warranted. This paper presents a linear-quadratic (LQ) control strategy that is robust and that allows control restructuring to be carried out in a natural fashion.

Linear-quadratic control theory inherently provides a number of useful features, such as direct treatment of multi-input/multi-output problems, easy transfer from continuous- to discrete-time descriptions, and guaranteed stability when certain criteria are satisfied.[1] However, the choice of quadratic state- and control-weighting matrices (Q and R) is not always obvious. This problem is of particular concern when a control system is to be reconfigured or restructured in real time following failures. Suppose that the chosen Q and R lead to desirable closed-loop characteristics for the nominal system (described by system and control matrices, F_1 and G_1). If failures cause the open-loop system to change (to F_2 and G_2), a controller designed with the same Q and R will have different closed-loop characteristics. Although the original cost function is, by definition, still optimized, the resulting response may no longer be adequate in a subjective sense. If similar system behavior is desired before and after the failure, it would be necessary in most cases to find new values of Q and R, which may be a nontrivial selection process.

Model following is an attractive candidate for the redesign process because the goal is to emulate the performance characteristics of a desirable model, with or without failures. This can be accomplished explicitly (by requiring the system outputs to follow those of the model in a least-squares sense) or implicitly (by minimizing a quadratic function of the actual and modeled state rates).[2,3] Explicit model following normally is less sensitive to parameter variations, but the model states must be generated, and feedback gains may be high for satisfactory performance. Implicit model following can be implemented more simply with lower gains, and the weighting matrices are directly affected by the difference in plant and model dynamics. Consequently, the effective Q and R of the LQ problem can be adjusted by known changes in plant dynamics so as to minimize the variation in closed-loop performance.

Model-following criteria were discussed by Erzberger,[2] who showed that explicit and implicit model following are equivalent provided certain conditions are satisfied. The synthesis of model-following systems was addressed by Tyler,[3] who analyzed feedback and feed-forward gains in terms of plant and model characteristics. Chan[4] examined the relative merits of explicit and implicit model-following controls and concluded that the former yielded smaller steady-state error when imperfect model following occurred. Asseo[5] formulated a type-one perfect model-following control law and found that it reduced sensitivity to parameter variations. Kreindler and Rothschild[6] also compared both model-following strategies and found that the implicit model-following method tracked the model better, especially during the initial transient. Broussard and Berry[7] considered a different formulation of implicit model following, in which feedback gains are determined from algebraic relationships. They showed that the approach is equivalent to pole-placement techniques. Kaufman[8] employed model-following concepts in the design of adaptive control systems. Digital implicit model following was studied in Ref. 9, and flight test results for a simple algebraic design method are presented in Ref. 10.

The subject of restructurable control is an area of ongoing research. Huber et al.[11] designed a self-repairing control system that utilized a control mixer concept to distribute control authority after a surface failure to remaining effectors. Os-

Presented as Paper 87-2312 at the AIAA Guidance, Navigation, and Control Conference, Monterey, CA, Aug. 17-19, 1987; received Dec. 1, 1987; revision received Dec. 23, 1988. Copyright © 1989 American Institute of Aeronautics and Astronautics, Inc. All rights reserved.

*Graduate Student, Department of Mechanical and Aerospace Engineering; currently with Grumman Corporation.

†Professor, Department of Mechanical and Aerospace Engineering. Associate Fellow AIAA.

troff and Hueschen[12] applied a command-generator-tracker/proportional-integral filter control law to a commercial airplane with control surface failures and under turbulence. Looze et al.[13] reported on an automatic redesign technique that restructures control such that a frequency-domain system performance metric is maximized.

In this paper, a proportional-integral implicit model-following (PIIMF) control law is outlined, and associated features are discussed. Some properties of the implicit model-following scheme and conditions for control reconfiguration, given changes in control and system matrices, are examined. PIIMF control is applied to an example problem based on the lateral-directional motions of a typical fighter aircraft. Automatic restructuring is illustrated for the changes that might be brought about by control actuator failures or aircraft structural and aerodynamic damage. The results are summarized in the conclusion.

Control Law Formulation

The system to be controlled, the desired dynamic model, and the cost function to be minimized are, respectively,

$$\dot{x} = Fx + Gu \tag{1}$$

$$\dot{x}_m = F_m x_m \tag{2}$$

$$J = \frac{1}{2}\int_0^\infty \left[(x - x_m)^T Q_1 (x - x_m)\right] dt \tag{3}$$

or

$$J = \frac{1}{2}\int_0^\infty \left[x^T Q x + 2x^T M u + u^T R u\right] dt \tag{4}$$

where x and u are actual state and control vectors of dimensions $n \times 1$ and $r \times 1$, respectively, x_m is an $n \times 1$ model state vector, F and G are system and control matrices for the system to be controlled, F_m is a model matrix, and Q_1 is a weighting matrix for the state-rate errors. Substituting Eqs. (1) and (2) into Eq. (3), and assuming that $x = x_m$, leads to Eq. (4), where $Q = (F - F_m)^T Q_1 (F - F_m)$, $M = (F - F_m)^T Q_1 G$, and $R = G^T Q_1 G$. From Eq. (4), we see that the PIIMF controller generates a cross-weighting matrix M. Considering optimality and pole-assignment requirements,[14,15] the presence of M is essential for achieving complete control over the system.

Equation (4) is simply a quadratic cost function of the original system [(Eq. (1)]. In addition to the inherent sensitivity to plant parameter variations,[5] we can improve the low-frequency disturbance-rejection characteristics of the control law and achieve zero steady-state error command response by adding proportional-integral (PI) compensation. Let y_d be the desired equilibrium outputs; then the perturbations around the corresponding set point (x_0, u_0) are

$$\Delta x(t) = x(t) - x_0 \tag{5}$$

$$\Delta u(t) = u(t) - u_0 \tag{6}$$

$$\Delta y(t) = y(t) - y_d \tag{7a}$$

$$= Hx(t) - y_d \tag{7b}$$

where H is the output matrix. The output error integral ξ is defined as

$$\xi = \xi(0) + \int_0^t \Delta y(\tau)\, d\tau \tag{8}$$

The state variable is augmented to include ξ,

$$\Delta \bar{x}_a(t) = \begin{bmatrix}\Delta x^T(t) & \xi^T(t)\end{bmatrix}^T = \begin{bmatrix}F & 0\\ H & 0\end{bmatrix}\Delta x_a(t) + \begin{bmatrix}G\\ 0\end{bmatrix}\Delta u(t) \tag{9}$$

and the cost function becomes

$$J = \frac{1}{2}\int_0^\infty \left[\Delta x_a^T Z \Delta x_a + 2\Delta x_a^T S \Delta u + \Delta u^T R \Delta u\right] dt \tag{10}$$

where

$$Z = \begin{bmatrix}Q & 0\\ 0 & Q_1\end{bmatrix}, \qquad S = \begin{bmatrix}M\\ 0\end{bmatrix}$$

and Q_1 is the weighting on ξ. The LQ control obtained by optimizing Eq. (10) is

$$\Delta u(t) = -C\Delta x_a(t) = -\begin{bmatrix}C_1 & C_2\end{bmatrix}\begin{bmatrix}\Delta x(t)\\ \xi(t)\end{bmatrix} \tag{11}$$

Expanding and rearranging Eq. (11), we get

$$u(t) = u_0 - C_1[x(t) - x_0] + C_2\left\{\xi(0) + \int_0^t [y(\tau) - y_d]\, d\tau\right\} \tag{12}$$

We can see that the second term on the right side provides proportional feedback of the error, whereas the third term provides the integral effect being sought. The dimension of the commanded input y_d is chosen to equal the dimension of u, reflecting the fact that the number of controllable states is equal to the number of the controls available.[1]

Properties of the Control Law

We first examine the effect of adding the integral loop. The steady-state Riccati equation resulting from optimizing Eq. (10) is

$$\begin{aligned}&\begin{bmatrix}P_{11} & P_{12}\\ P_{12}^T & P_{22}\end{bmatrix}\left[\begin{pmatrix}F & 0\\ H & 0\end{pmatrix} - \begin{bmatrix}G\\ 0\end{bmatrix}R^{-1}\begin{bmatrix}M^T & 0\end{bmatrix}\right]\\ &+ \left[\begin{pmatrix}F & 0\\ H & 0\end{pmatrix} - \begin{bmatrix}G\\ 0\end{bmatrix}R^{-1}\begin{bmatrix}M^T & 0\end{bmatrix}\right]^T\begin{bmatrix}P_{11} & P_{12}\\ P_{12}^T & P_{22}\end{bmatrix}\\ &- \begin{bmatrix}P_{11} & P_{12}\\ P_{12}^T & P_{22}\end{bmatrix}\begin{bmatrix}G\\ 0\end{bmatrix}R^{-1}\begin{bmatrix}G^T & 0\end{bmatrix}\begin{bmatrix}P_{11} & P_{12}\\ P_{12}^T & P_{22}\end{bmatrix}\\ &+ \begin{pmatrix}Q & 0\\ 0 & Q_1\end{pmatrix} - \begin{bmatrix}M\\ 0\end{bmatrix}R^{-1}\begin{bmatrix}M^T & 0\end{bmatrix} = 0\end{aligned} \tag{13}$$

Expanding along each submatrix, we get

$$\begin{aligned}&P_{11}(F - GR^{-1}M^T) + (F - GR^{-1}M^T)^T P_{11} - P_{11}GR^{-1}G^T P_{11}\\ &+ Q - MR^{-1}M^T + P_{12}H + H^T P_{12}^T = 0\end{aligned} \tag{14}$$

$$P_{12}^T(F - GR^{-1}M^T) + P_{22}H - P_{12}^T GR^{-1}G^T P_{11} = 0 \tag{15}$$

$$-P_{12}^T GR^{-1}G^T P_{12} + Q_1 = 0 \tag{16}$$

The Riccati equation obtained by optimizing the implicit model-following strategy alone [Eq. (4)] is:

$$\begin{aligned}&P(F - GR^{-1}M^T) + (F - GR^{-1}M^T)^T P\\ &- PGR^{-1}G^T P + Q - MR^{-1}M^T = 0\end{aligned} \tag{17}$$

Therefore, the effect of the integration loop is to add the terms $P_{12}H + H^T P_{12}^T$. Since from Eq. (16) we see that P_{12} is determined by G and Q_1, and it is independent of P_{11} and the model, these extra terms constitute a shift or "bias" in state-weighting Q.

We now proceed to the analysis of the implicit model-following strategy, making comments on the integrator effect whenever necessary. If G were invertible (which would require the number of controls to be equal to the number of states) and Q_1 were diagonal, then from the definitions of Q, M, and R,

$$(F - GR^{-1}M^T) = F - GR^{-1}\left[(F - F_m)^T Q_i G\right]^T \quad (18a)$$

$$= F - GR^{-1}G^T Q_i^T (F - F_m) \quad (18b)$$

$$= F - G(G^T Q_i G)^{-1} G^T Q_i^T (F - F_m) \quad (18c)$$

$$= F - GG^{-1}Q_i^{-1}G^{-T}G^T Q_i (F - F_m) \quad (18d)$$

$$= F - F + F_m \quad (18e)$$

$$= F_m \quad (18f)$$

$$GR^{-1}G^T = GR^{-1}G^T Q_i Q_i^{-1} = Q_i^{-1} \quad (19)$$

$$MR^{-1}M^T = (F - F_M)^T Q_i G(G^T Q_i G)^{-1} G^T Q_i (F - F_m) \quad (20a)$$

$$= (F - F_m)^T Q_i G R^{-1} G^T Q_i (F - F_m) \quad (20b)$$

$$= (F - F_m)^T Q_i (F - F_m) = Q \quad (20c)$$

We have intentionally isolated the term $GR^{-1}G^T Q_i$, which is equal to the identity matrix in this case; its significance will become apparent later. Inserting the last three relations into Eq. (17), we get

$$PF_m + F_m^T P - PQ_i^{-1}P = 0 \quad (21)$$

This can be construed as a Riccati equation resulting from optimizing a model of the form $\dot{x}_m = F_m x_m + u_m$ for a cost function consisting of no state weighting and Q_i control weighting. Assuming F_m is stable, the unique solution to this equation is $P = 0$, which means $u_m = 0$; thus, the model system is recovered. The feedback gain C and closed-loop system F_c of the original system become

$$C = R^{-1}(G^T P + M^T) \quad (22a)$$

$$= G^{-1}Q_i^{-1}G^{-T}\left[G^T Q_i (F - F_m)\right] \quad (22b)$$

$$= G^{-1}(F - F_m) \quad (22c)$$

$$F_c = F - GC = F_m \quad (23)$$

Therefore, the gain would be proportional to the available control authority and to the difference between plant and model dynamics, whereas the closed-loop dynamics would simply become those of the model. Not surprisingly, Eq. (22) is identical to Erzberger's perfect model-following control law.[2] When the integral loop is present, P is no longer zero, and we have

$$C = R^{-1}\left\{\begin{bmatrix} G^T & 0 \end{bmatrix}\begin{bmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} + \begin{bmatrix} M^T & 0 \end{bmatrix}\right\} \quad (24a)$$

$$C = \begin{bmatrix} R^{-1}G^T P_{11} + G^{-1}(F - F_m) & R^{-1}G^T P_{12} \end{bmatrix} \quad (24b)$$

$$F_c = \begin{bmatrix} F & 0 \\ H & 0 \end{bmatrix} - \begin{bmatrix} G \\ 0 \end{bmatrix}\begin{bmatrix} R^{-1}G^T P_{11} \\ + G^{-1}(F - F_m) & R^{-1}G^T P_{12} \end{bmatrix} \quad (25a)$$

$$F_c = \begin{bmatrix} F_m - Q_i^{-1}P_{11} & -Q_i^{-1}P_{12} \\ H & 0 \end{bmatrix} \quad (25b)$$

Hence, the gain is no longer given by Eq. (22), and the closed-loop system characteristics, as indicated by the damping and transient behavior, will be different from that of the model. Both attributes are functions of P_{12}, which is determined by the weighting matrices Q_1 and Q_i through Eq. (16).

Consider now the term $GR^{-1}G^T Q_i$ in the more typical case with fewer controls than states,

$$GR^{-1}G^T Q_i = G(G^T Q_i G)^{-1} G^T Q_i \quad (26a)$$

$$= GG^{\#1} \quad (26b)$$

where $G^{\#1}$ is the weighted left pseudo-inverse[3] of G. When the structure of the control matrix G (of rank r) is such that it contains an $r \times r$ submatrix much larger (in the two-norm sense) than the remaining submatrices, and Q_i does not change the inequality by its choice of weighting, it can be shown[16] that

$$GG^{\#1} \approx I_w \quad (27)$$

where I_w is a matrix of dimension $n \times n$ with r ones in the diagonals and zeros everywhere else. Physically, the requirement is that r controls have a stronger effect on r state components (e.g., the angular rates) and weaker effect on the remaining states (e.g., the angles). Moreover, let I_w^* be the "complement" of I_w; that is, $I_w^* = (I_n - I_w)$, where I_n is the $n \times n$ identity matrix; then Eqs. (18) and (20) become

$$F - GR^{-1}M^T = F - GR^{-1}G^T Q_i (F - F_m) \quad (28a)$$

$$\approx F - I_w (F - F_m) \quad (28b)$$

$$= I_w F_m + I_w^* F = F^* \quad (28c)$$

$$MR^{-1}M^T = (F - F_m)^T I_w Q_i (F - F_m) \quad (29)$$

If we let $Q' = Q - (F - F_m)^T I_w Q_i (F - F_m)$, the Riccati equation [Eq. (13)] can be approximated as

$$PF^* + F^{*T}P - PI_w Q_i^{-1}P + Q' = 0 \quad (30)$$

where F^* is defined by Eq. (28). The gain and closed-loop system matrices are

$$C = G^{\#1}[Q_i^{-1}P + (F - F_m)] \quad (31)$$

$$F_c = I_w F_m + I_w^* F - I_w Q_i^{-1}P \quad (32a)$$

$$= F^* - I_w Q_i^{-1}P \quad (32b)$$

$$= I_w^*(F - F_m) + F_m - I_w Q_i^{-1}P \quad (32c)$$

We note that this development bears resemblance to the perfect model-following condition as stated by Erzberger[2]

$$(GG^{\#} - I)(F - F_m) = 0 \quad (33)$$

From Eq. (28) [or Eq. (18)], we have

$$GR^{-1}M^T = GR^{-1}G^T Q_i (F - F_m) \approx I_w (F - F_m) \quad (34)$$

or

$$0 \approx (GR^{-1}G^T Q_i - I_w)(F - F_m) \quad (35a)$$

$$\approx (GG^{\#1} - I_w)(F - F_m) \quad (35b)$$

The closer $GR^{-1}G^T Q_i$ is to the quasi-identity matrix I_w, the better the plant behaves like the model; the closeness also affects the control restructuring strategy to be described later. As will be shown, Eq. (35a) is one of the control reconfiguration criteria. The effect of the integrator is similar to the case where G has full rank. This can be seen by computing P_{12} from Eq. (16) using $I_w Q_i^{-1}$ in place of $GR^{-1}G^T$ and substituting the result into Eq. (13).

Effects of Control Reconfiguration

Reconfiguration with Variations in Control Effect

One case of interest is the degradation of a control effector. This type of failure can be approximated by reducing a column of the control matrix by a common factor. This is equivalent to an elementary row operation of matrix algebra. Let $E = \text{diag}[1 \quad 1 \cdots \alpha \cdots 1]$, $\alpha > 0$, where α is the factor by which the

effectiveness is reduced. If G' ($=GE$) is the new control matrix, the weighting matrices and Eqs. (18–20) then become

$$R' = G'^T Q_y G' = E^T G^T Q_y G E = E^T R E \tag{36}$$

$$M' = (F - F_m) Q_y G' = M E \tag{37}$$

$$(F - G' R'^{-1} M'^T) = (F - G E E^{-1} R^{-1} E^{-T} E^T M) \tag{38a}$$

$$= (F - G R^{-1} M^T) \tag{38b}$$

$$G' R'^{-1} G' = G E E^{-1} R^{-1} E^{-T} E^T G = G R^{-1} G^T \tag{39}$$

$$M' R'^{-1} M'^T = M E E^{-1} R^{-1} E^{-T} E^T M = M R^{-1} M^T \tag{40}$$

Hence, there is no change in the Riccati equation (Eq. 13 or Eq. 17). As expected, the gain is altered since R is decreased so as to allow additional control activity, but the closed-loop system remains unchanged:

$$C' = R'^{-1}(G'^T P + M'^T) \tag{41a}$$

$$= E^{-1} R^{-1} E^{-T} (E^T G^T P + E^T M) = E^{-1} C \tag{41b}$$

$$F'_c = F - G' C' = F - G E E^{-1} C = F - G C = F_c \tag{42}$$

Since the integral effect enters only through P_{12} [Eq. (16)], which is unaffected by this type of failure, the analysis holds. The control restructuring process is now clear. When the gains are recalculated with G', the original system performance is restored. Note that these results are independent of the invertibility of G.

If $\alpha = 0$, reconfiguration is addressed simply by reducing the control vector dimension and adjusting G, R, and M accordingly. In this situation, it is important to insure that controllability has been retained with the reduced-dimension controller.

The effects of single- and multiple-element changes are more difficult to analyze since they no longer can be represented as elementary matrix operations. However, Eq. (26a) [or equivalently Eq. (35a)] can still be computed using G' ($= G + \delta G$) in place of G. As long as the equality [Eq. (27)] (approximately) holds, the controls can force the plant to behave like the model. Otherwise, the control matrix must have undergone substantial changes, such that the original performance cannot be approximated by the implicit model-following strategy.

Reconfiguration with Variations in System Dynamics

System dynamics could change, for example, as a consequence of structural damage. We first examine the reconfiguration when a row in F changes by a common factor. Row changes are equivalent to premultiplication by the elementary matrix E; that is, $F' = EF$. If G is invertible, then the previous analysis holds, and the performance can be recovered by simply recalculating the gain using F' instead of F. For nonsquare G, when the changes are along the nonzero rows of I_m, Q and M are modified such that the solution of the Riccati equation [Eq. (13)] is not significantly affected. In this case, the gain and the closed-loop system are

$$C = R^{-1}(G^T P + M'^T) \tag{43a}$$

$$= G^{\#L}[Q_y^{-1} P + (F' - F_m)] \tag{43b}$$

$$F'_c = F' - GC = EF - G G^{\#L} Q_y^{-1} P - G G^{\#L}(EF - F_m) \tag{44a}$$

$$= I_m F_m - I_m^R E F - I_m Q_y^{-1} P \tag{44b}$$

$$= I_m^R (F' - F_m) + F_m - I_m Q_y^{-1} P = F_c \tag{44c}$$

Thus, the gain is altered, but the closed-loop system remains essentially the same. As can be seen from Eqs. (32) and (44), the reconfiguration may not recover the performance if F is changed along the zero rows of I_m; that is, the following relation must hold:

$$I_m^R (F' - F_m) = I_m^R (F - F_m) \tag{45}$$

This equation constitutes the other reconfiguration criterion to be applied in the case of system dynamic changes. Note that since P_{12} does not depend on F, changes in the system dynamics have no effect on the integral loop.

Failures are unlikely to cause alterations in the manner described. Single- or multiple-element failures by different factors are expected. If the reconfiguration criteria are satisfied, then the performance will be recovered when the gains are recalculated using F' ($= F + \delta F$).

We point out here that reconfiguration is possible with $GR^{-1}G^T Q_y$ not close to I_m, although acceptable performance may require it. If the expression exhibits a nonidentity matrix structure, then the model is not well followed; consequently, the performance of the reconfigured system would not be near its nominal values. Moreover, since P_{12} does not depend on F, changes in the system dynamics have no effect on the integral loop; therefore the analysis carries over to the PIIMF control law. This and other observations are verified numerically in the next section.

Simulation Results

The analysis described in the previous sections is applied to the lateral-axis model of the F-4 aircraft. The state vector consists of roll rate p (rad/s), yaw rate r (rad/s), sideslip β (rad), and roll angle ϕ (rad); i.e., $x = [p\ r\ \beta\ \phi]^T$. The controls are rudder δr(in.) and aileron δa(in.); i.e., $u = [\delta r\ \ \delta a]^T$. The system, control, and model matrices are[4]

$$F = \begin{bmatrix} -1.768 & 0.415 & -14.25 & 0.0 \\ -0.0007 & -0.3831 & 6.038 & 0.0 \\ 0.0016 & -0.9975 & -0.1551 & 0.0586 \\ 1.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \tag{46}$$

$$G = \begin{bmatrix} 1.744 & 8.952 \\ -2.92 & -0.3075 \\ 0.0243 & -0.0036 \\ 0.0 & 0.0 \end{bmatrix} \tag{47}$$

$$F_m = \begin{bmatrix} -4.0 & 0.865 & -10.0 & 0.0 \\ 0.04 & -0.507 & 5.87 & 0.0 \\ 0.0016 & -1.0 & -0.743 & 0.0586 \\ 1.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \tag{48}$$

$$Q_y = \text{diag}[0.01 \quad 0.01 \quad 0.01 \quad 0.01] \tag{49}$$

$$Q_t = \text{diag}[1.0 \quad 1.0] \tag{50}$$

The speed of response is determined by the ratio of Q_y to Q_t. The weighting matrices as shown are chosen to produce reasonable tracking; no extensive search for other combinations was conducted.

For this system, Eq. (26a) is found to be

$$GR^{-1}G^TQ_c = \begin{bmatrix} 1.0 & -6.34\times10^{-6} & -0.0007 & 0.0 \\ -6.256\times10^{-6} & 1.0 & -0.00874 & 0.0 \\ -0.0007 & -0.00874 & 7.7\times10^{-5} & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \tag{51}$$

This can be seen to be approximately equal to I_{22}. Therefore, two model states (p and r) can be followed, and the reconfiguration is expected to work well for changes that affect primarily these two variables.

The time histories of the rudder, aileron, sideslip, and roll angle subject to a 1 rad roll-angle command and 0.5 rad sideslip (a rather unusual maneuver) are used as performance evaluations. Several failure cases are examined here.

Case 1

Aileron effectiveness is reduced by half; that is, the second column of G is multiplied by 0.5. This failure is equivalent to a loss of aileron surface area that decreases the available rolling control moment. The failed response is shown as a broken curve in Fig. 1. The reconfigured system response is shown as the bold-face curve. Notice that the yaw and roll angle responses of the failed and unfailed systems are indistinguishable, whereas the aileron exhibits higher activity with no visible change in rudder motion. Corresponding gains C_1 and C_2 for this and other cases are listed in Table 1.

Case 2

The first row of the F matrix and the second column of G are reduced by half. This corresponds to an overall decrease in

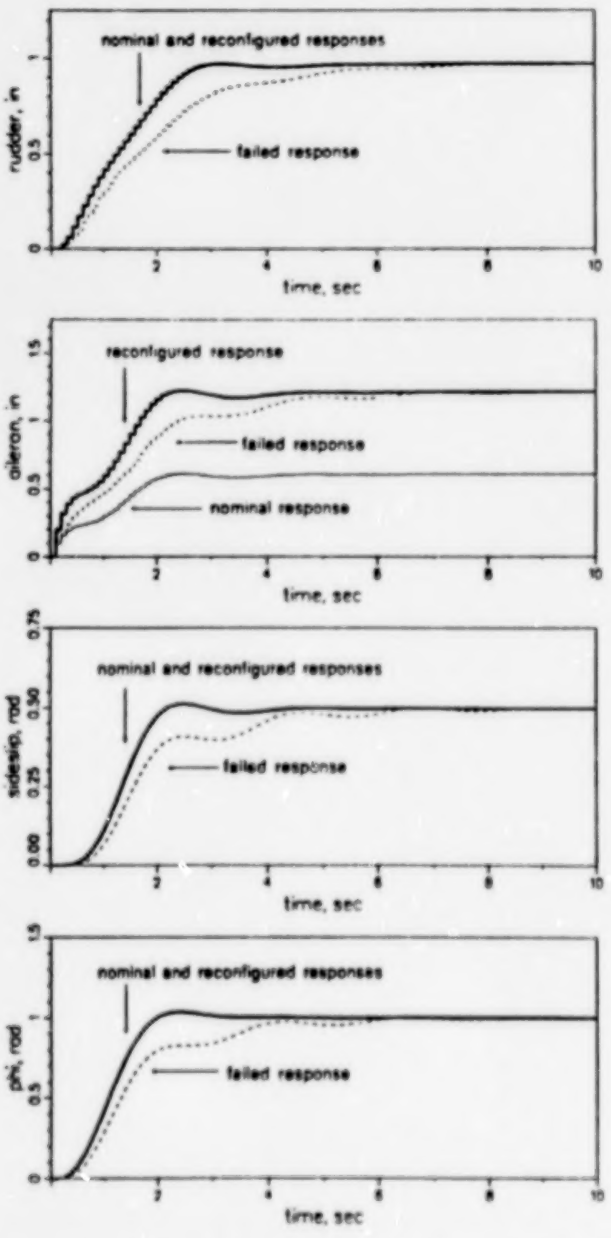


Fig. 1 Control and state responses of the failure case 1.

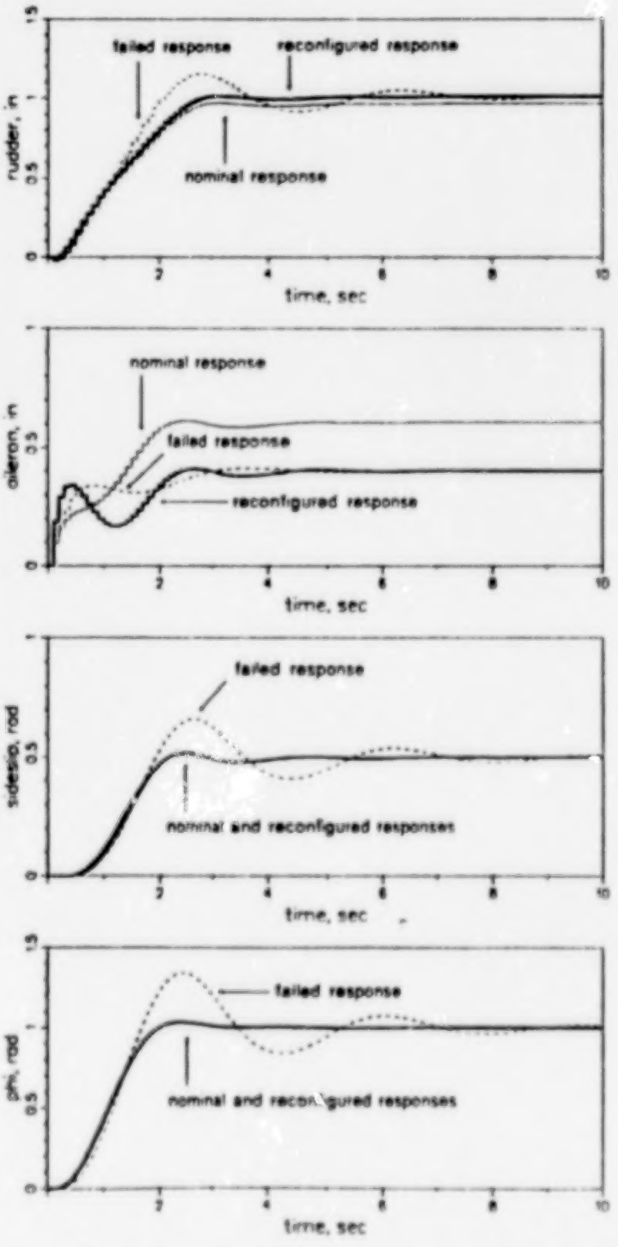


Fig. 2 Control and state responses of the failure case 2.

Table 1 Control gains for sample scenarios

Scenario	C_1				C_2	
No failure case	−0.3892	−1.07871	1.99	−1.9472	2.9054	−1.95
	0.517	0.272	−1.1484	1.36625	0.024	1.3282
Failure case 1	−0.3892	−1.0787	1.99	−1.9472	2.9054	−1.95
	1.034	0.544	−2.3	2.7325	0.048	2.656
Failure case 2	−0.4	−1.077	1.909	−1.948	2.905	−1.95
	1.235	0.497	−0.6733	2.7325	0.0485	2.656
Failure case 3	−22.181	−32.4	108.94	−122.3106	147.287	−152.09
	−3.741	6.331	−21.806	23.696	−27.914	30.43
Failure case 4	−0.312	−0.795	4.16	−1.477	3.28	−1.217
	0.53	0.1893	−1.823	1.417	−0.286	1.297

aerodynamic forces and control effectiveness that contribute to the roll rate, as might result from battle damage to the wings. The failed and reconfigured responses are shown in Fig. 2. Note that the control activity is less in the reconfigured case since the ratio of control to the aerodynamic effectiveness is larger than that of the nonfailed system.

Case 3

The (2,1) element of G (proportional to the aileron yaw effectiveness $C_{n_{\delta_a}}$) is reduced from −2.92 to −0.1 (a significant change). This cross-coupling alteration may be one of the many other changes induced by damages; it is isolated here to illustrate its effect. The responses are shown in Fig. 3. For this failure, $GR^{-1}G^TQ_i$ is:

$$GR^{-1}G^TQ_i = \begin{bmatrix} 1.0 & -0.0003 & -0.0012 & 0.0 \\ -0.0003 & 0.934 & -0.2484 & 0.0 \\ -0.0012 & -0.2484 & 0.0661 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \quad (52)$$

Because the expression departs from its quasi-identity-matrix structure, the reconfiguration is unable to recover the original response.

Case 4

The third row of the system matrix is reduced by a factor of 0.25. The responses are shown in Fig. 4. For this case, Eq. (45) no longer holds, and we expect the reconfiguration to perform poorly. The integral compensator still forces zero steady-state error, albeit at a much longer time constant. Note that this is not a physically realizable case since it involves changing a pure integration factor. It is used here to illustrate the importance of reconfiguration criterion 2. A more realistic scenario and additional cases can be found in Ref. 16.

A few additional observations are due. $GR^{-1}G^TQ_i$ remained close to I_{22} despite the changes in the first two failure cases where control restructuring performed well. When aileron effectiveness was reduced, the reconfigured system simply increased the aileron motions with no apparent change in rudder activity. This is because the rudder is an ineffective surface for roll control at this flight condition. In a more cross-coupled system, the control activities would be more evenly reallocated. To reduce control motions, control-limiting criteria can be incorporated into the cost function.

Similarly, model uncertainty can be taken into account by loop-shaping techniques at the outset; however, robustness properties are altered by control reconfiguration. The integral effect is largely unaffected by the failure, even though the ξ response is different due to different plant dynamics. No updating of Q_i is necessary if the ratio of the integration rate and error feedback is to be preserved. Complete control failures were not illustrated, but it can be shown (see Ref. 1) that system performance cannot be totally restored, and the system behavior will depend on how well the remainder of the controls can achieve the model following.

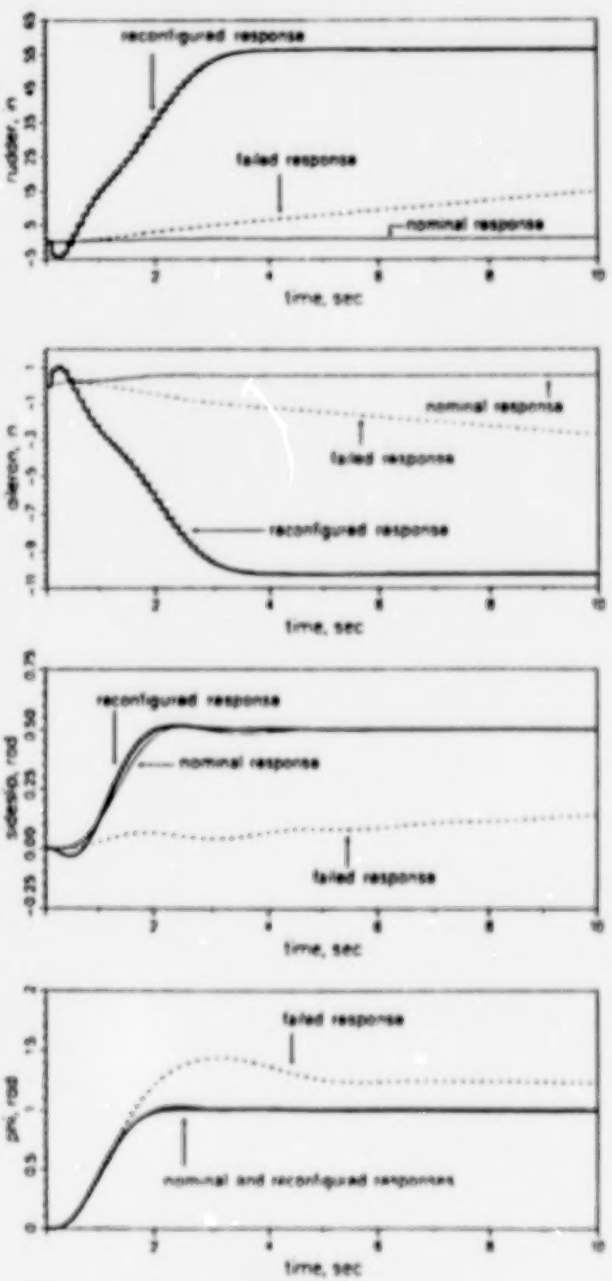


Fig. 3 Control and state responses of the failure case 3.

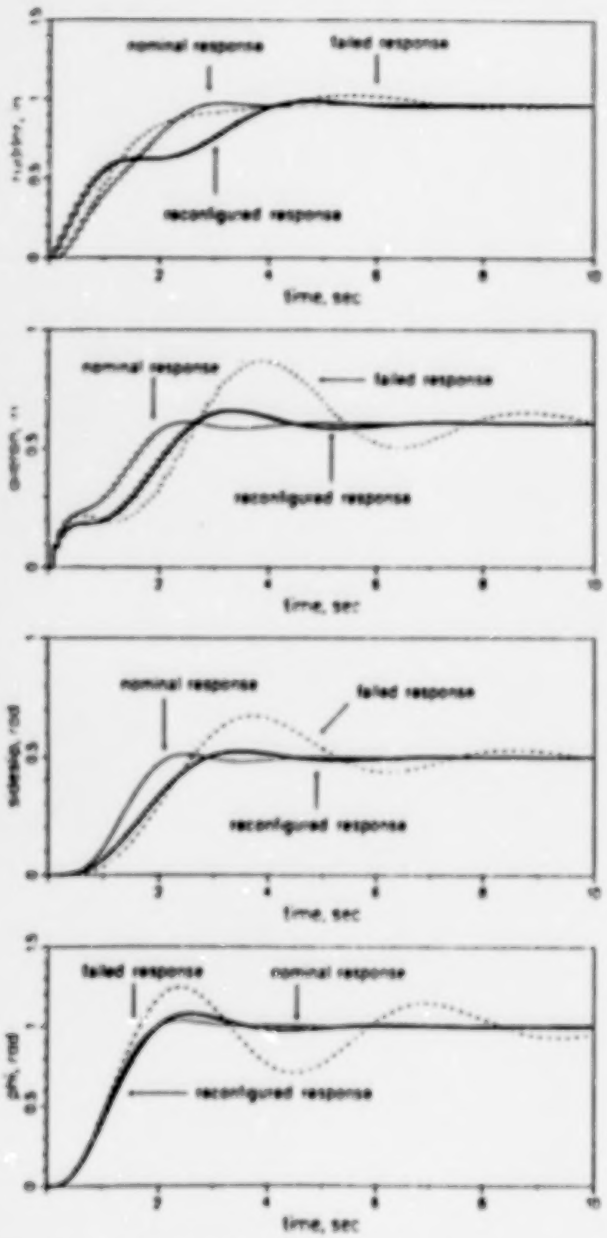


Fig. 4 Control and state responses of the failure case 4.

Conclusions

An implicit model-following control strategy for restructuring a control system following control-effector failures and alterations in system dynamics has been presented. Examination of the inherent properties of the control scheme led to the formulation of two reconfiguration criteria that, if satisfied, guarantee the performance recovery of a failed system when the feedback gains are recalculated using the new dynamics. Care must be exercised in applying the strategy since, in some cases, higher demand is placed on the remaining control authority. However, when properly used, the proportional-integral implicit model-following control law is a good candidate for control reconfiguration.

The implicit model-following strategy reconfigures the system by rearranging the weighting matrices to preserve closed-loop characteristics. This is analogous to a pole-placement technique that assigns the closed-loop poles to the same locations before and after the failure. However, implicit model following does more than positioning eigenvalues and matching desired eigenvectors; it also addresses command response, which is not totally specified by eigenvalues and eigenvectors. The advantages of the model-following scheme are that, following partial control failure or changes in the system dynamic characteristics, it attempts to maintain the desired eigenstructure and command response, and at the same time, it allows tradeoffs between control activity and system response.

Acknowledgment

This work was performed under Contract DAAG29-84-K-0048 for the U.S. Army Research Office.

References

[1]Stengel, R., *Stochastic Optimal Control: Theory and Application*, Wiley, New York, 1986.

[2]Erzberger, H., "On the Use of Algebraic Methods in the Analysis and Design of Model Following Control System," NASA TN D-4663, 1968.

[3]Tyler, J., Jr., "The Characteristics of Model-Following Systems as Synthesized by Optimal Control," *IEEE Transactions on Automatic Control*, Vol. AC-15, June 1970, pp. 326-333.

[4]Chan, Y., "Perfect Model Following with a Real Model," *Joint Automatic Control Conference Preprints*, Inst. of Electrical and Electronics Engineers, New York, 1973, pp. 287-291.

[5]Asseo, S., "Application of Optimal Control to Perfect Model Following," *Journal of Aircraft*, Vol. 7, No. 4, 1970, pp. 308-313.

[6]Kreindler, E., and Rothschild, D., "Model-Following in Linear Quadratic Optimization," *AIAA Journal*, Vol. 14, No. 7, 1976, pp. 835-842.

[7]Broussard, J., and Berry, P., "The Relationship Between Implicit Model Following and Eigenvalue Eigenvector Placement," *IEEE Transactions on Automatic Control*, Vol. AC-25, June 1980, pp. 591-594.

[8]Kaufman, H., "Research in Digital Adaptive Flight Controllers," NASA CR-2684, May 1976.

[9]Stengel, R., "Digital Flight Control Design Using Implicit Model Following," AIAA Paper 73-844, Aug. 1973.

[10]Grunwald, S., and Stengel, R., "Design and Flight Testing of Digital Direct Side-Force Control Laws," *Journal of Guidance, Control, and Dynamics*, Vol. 8, No. 2, 1985, pp. 188-193.

[11]Huber, R. R., *et al.*, "Self-Repairing Digital Flight Control System Study," Air Force Flight Dynamics Lab., Wright-Patterson AFB, OH, TR (draft), Pt. 1, July 1984.

[12]Ostroff, A., and Hueschen, R., "Investigation of Control Law Reconfigurations to Accommodate a Control Element Failure," *Proceedings of the American Control Conference*, Inst. of Electrical and Electronics Engineers, New York, June 1984, pp. 1746-1754.

[13]Looze, D., Weiss, J., Eterno, J., and Barrett, N., "An Automatic Redesign Approach for Restructurable Control System," *Proceedings of the National Aeronautic Electrical Conference*, Inst. of Electrical and Electronics Engineers, New York, May 1985, pp. 570-577.

[14]Kreindler, E., and Jameson, A., "Optimality of Linear Control Systems," *IEEE Transactions on Automatic Control*, Vol. AC-17, 1972, pp. 329-351.

[15]Hedrick, J., and Youcef-Toumi, K., "Exact Multi-Input Pole Placement by Linear Quadratic Synthesis," *Proceedings of the American Control Conference*, Inst. of Electrical and Electronics Engineers, New York, June 1983, pp. 905-910.

[16]Huang, C., "A Methodology for Knowledge-Based Restructurable Control to Accommodate System Failures," Ph.D. Thesis, Princeton Univ., Princeton, NJ, 1989.

STOCHASTIC STABILITY AND PERFORMANCE ROBUSTNESS OF LINEAR MULTIVARIABLE SYSTEMS

Laura E. Ryan* and Robert F. Stengel**

Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, N.J.

ABSTRACT

Stochastic robustness, a simple technique used to estimate the robustness of linear, time-invariant systems, is applied to a single-link robot arm control system. Concepts behind stochastic *stability* robustness are extended to systems with estimators and to stochastic *performance* robustness. Stochastic performance robustness measures based on classical design specifications are introduced, and the relationship between stochastic robustness measures and control system design parameters is discussed. The application of stochastic performance robustness, and the relationship between performance objectives and design parameters are demonstrated by means of the example. The results prove stochastic robustness to be a good overall robustness analysis method that can relate robustness characteristics to control system design parameters.

INTRODUCTION

Standard linear control system design techniques rely on accurate models of the system to be controlled. Because models are never perfect, robustness analysis is necessary to determine the possibility of instability or inadequate performance in the face of uncertainty. Robustness to these uncertainties, parametric or unstructured, is normally treated deterministically and often without regard to possible physical variations in the system. Consequently, overconservative control system designs, or designs that are insufficiently robust in the face of real world uncertainties are a danger.

Stochastic robustness, a simple technique to determine the robustness of linear, time-invariant systems by Monte Carlo methods was introduced in [1] and presented in detail in [2,3]. These references described stochastic stability robustness analysis and introduced the probability of instability as a scalar measure of stability robustness. Confidence intervals for the scalar probability of instability were presented, and the stochastic root locus, or probability density of the closed-loop eigenvalues, has shown to portray robustness properties graphically. Because it is a statistical measure of robustness, and because it directly uses knowledge of the statistics of the physical parameter variations, stochastic robustness is inherently intuitive and precise. The physical meaning behind the probability of instability is apparent, and overconservative or insufficiently robust designs can be avoided. Applications of stochastic robustness to analyzing full-state feedback aircraft control systems were described in [4]. The results presented there illustrated the use of stochastic stability robustness techniques in comparing control system designs and in including finite-dimensional uncertain dynamics.

Concepts behind stochastic stability robustness can be easily extended to provide insight into control system design for performance. Design specifications such as rise time, overshoot, settling time, dead time, and steady-state error are normally used as indicators of adequate performance and lend themselves to the same kind of analysis as described above. Concepts of stochastic stability robustness analysis can be applied to these criteria giving probabilistic bounds on individual scalar performance criteria. Metrics resulting from stability and performance robustness can be related to controller parameters, thus providing a foundation for design tradeoffs and optimization. Details of these extensions and uses of stochastic robustness are described in the following and are illustrated by means of an example.

STOCHASTIC STABILITY ROBUSTNESS

Stochastic stability robustness of a linear, time-invariant (LTI) system was described in [3] and is summarized here. Consider a LTI system subject to constant-coefficient control:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{p})\mathbf{x}(t) + \mathbf{G}(\mathbf{p})\mathbf{u}(t) \tag{1}$$

$$\mathbf{y}(t) = \mathbf{H}(\mathbf{p})\mathbf{x}(t) \tag{2}$$

$$\mathbf{u}(t) = \mathbf{u}_C(t) - \mathbf{C}\mathbf{H}(\mathbf{p})\mathbf{x}(t) \tag{3}$$

$\mathbf{x}(t)$, $\mathbf{u}(t)$, $\mathbf{y}(t)$, and $\mathbf{p}$ are state, control, output, and parameter vectors of dimension n, m, q, and r, respectively, accompanied by conformable dynamic, control, and output matrices that may be arbitrary functions of $\mathbf{p}$. $\mathbf{u}_C(t)$ is a command input vector, and, for simplicity, the (m x n) control gain matrix $\mathbf{C}$ is assumed to be known without error. The n eigenvalues, $\lambda_i = \sigma_i + j\omega_i$, $i = 1$ *to* n, of the matrix $[\mathbf{F}(\mathbf{p}) - \mathbf{G}(\mathbf{p})\mathbf{C}\mathbf{H}(\mathbf{p})]$ determine closed-loop stability. The control gain matrix $\mathbf{C}$ is designed using some nominal or "mean" value of the dynamic model, denoted $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H}$, that represents $\mathbf{F}(\mathbf{p})$, $\mathbf{G}(\mathbf{p})$, $\mathbf{H}(\mathbf{p})$ evaluated at the nominal parameter vector. The actual system has an unknown description, denoted $\mathbf{F}_\mathbf{A}$, $\mathbf{G}_\mathbf{A}$, and $\mathbf{H}_\mathbf{A}$ that depends on the actual (unknown) value of the parameter vector $\mathbf{p}$. Environment, variations in the nominal state, system failures, parameter estimation errors, wear, and manufacturing differences all can contribute to mismatch between the actual system and that used to design the controller. The parameter vector $\mathbf{p}$ is assumed to have a known or estimated probability density function, denoted pr($\mathbf{p}$), that expresses the statistics of parametric uncertainty due to the above factors.

System stability requires that no eigenvalues have positive real parts. While the relationship between parameters and eigenvalues is complicated, estimating the probability of instability ($\hat{\mathrm{P}}$) of a closed-loop system from repeated eigenvalue calculation is a straightforward task. Using *Monte Carlo evaluation*, the closed-loop eigenvalues are evaluated J times with each element of $\mathbf{p}_j$, $j = 1$ to J, specified by a random-number generator whose individual outputs are shaped by pr($\mathbf{p}$). The probability-of-instability estimate becomes increasingly precise as J becomes large. Then,

$$\Pr(\text{stability}) = \lim_{J \to \infty} \frac{N(\sigma_{\max} \le 0)}{J} \tag{4}$$

and

$$\Pr(\text{instability}) = \mathrm{P} = 1 - \Pr(\text{stability}) \tag{5}$$

---

* Graduate student
** Professor

Presented at the 1990 American Control Conference, San Diego, California.

$N(.)$ is the number of cases for which all elements of σ, the vector of the real parts of the closed-loop eigenvalues, are less than or equal to zero, that is, for which $\sigma_{max} \le 0$, where σ_{max} is the maximum real eigenvalue component in σ. For less than an infinite number of evaluations, the resulting Monte Carlo evaluation is an estimate, denoted $\hat{P}$.

Because P is a *binomial* variable (i.e., the outcome of each Monte Carlo evaluation takes on one of two values: *stable* or *unstable*) confidence intervals are calculated using the binomial test, where lower (L) and upper (U) intervals satisfy [5]

$$\Pr(X \le x\text{-}1) = \sum_{j=0}^{x-1} (J, j)\, L^j (1 - L)^{J-j} = 1 - \frac{\alpha}{2} \qquad (6$$

$$\Pr(X \le x) = \sum_{j=0}^{x} (J, j)\, U^j (1 - U)^{J-j} = \frac{\alpha}{2} \qquad (7$$

X is the actual number of unstable cases after J evaluations ($X = J\hat{P}$) and (J, x) is the binomial coefficient, $\frac{J!}{x!(J-x)!}$. Explicit approximations of the binomial test [6, 7] avoid an iterative solution of Eq. 6 and 7 for (L,U) and are accurate to within 0.1%. Confidence intervals for the unknown, true probability of instability are presented in more detail in [3].

Stochastic Stability Robustness of Systems with Estimators

Stochastic stability robustness analysis is easily extended to systems incorporating dynamic state estimators. Using $\mathbf{F_A}$, $\mathbf{G_A}$, and $\mathbf{H_A}$ as the actual system matrices and **F**, **G**, and **H** as the design system, state and estimator equations are [8]

$$\dot{\mathbf{x}}(t) = \mathbf{F_A}(\mathbf{p})\mathbf{x}(t) + \mathbf{G_A}(\mathbf{p})\mathbf{u}(t) + \mathbf{L_A}(\mathbf{p})\mathbf{w}(t) \qquad (8$$

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{F}\hat{\mathbf{x}}(t) + \mathbf{G}\mathbf{u}(t) + \mathbf{K}[\mathbf{z}(t) - \mathbf{H}\hat{\mathbf{x}}(t)] \qquad (9$$

$$\mathbf{u}(t) = \mathbf{u_c}(t) - \mathbf{C}\hat{\mathbf{x}}(t) \qquad (10$$

where $\hat{\mathbf{x}}(t)$ is the estimated state, **L** is the disturbance input matrix, and **K** is the estimator gain matrix. The measurement is taken through the actual output matrix and can include measurement noise:

$$\mathbf{z}(t) = \mathbf{H_A}(\mathbf{p})\mathbf{x}(t) + \mathbf{n}(t) \qquad (11$$

Stability of the full dynamic compensater is determined by the combined system. Substituting for the control in Eq. 8 and 9, the eigenvalues of the closed-loop system matrix $\mathbf{F_{cl}}$ must be in the left-half plane for stability:

$$\mathbf{F_{cl}} = \begin{bmatrix} \mathbf{F_A} & -\mathbf{G_A}\mathbf{C} \\ \mathbf{K}\mathbf{H_A} & \mathbf{F} - \mathbf{G}\mathbf{C} - \mathbf{K}\mathbf{H} \end{bmatrix} \qquad (12$$

In terms of the state ($\mathbf{x}$) and error dynamics ($\mathbf{x} - \hat{\mathbf{x}}$), the coupling effect of mismatch on the closed-loop system is directly apparent [8]

$$\mathbf{F_{cl}} = \begin{bmatrix} \mathbf{F_A} - \mathbf{G_A}\mathbf{C} & -\mathbf{G_A}\mathbf{C} \\ (\mathbf{F}-\mathbf{F_A}) - (\mathbf{G}-\mathbf{G_A}) - \mathbf{K}(\mathbf{H}-\mathbf{H_A}) & \mathbf{F} - (\mathbf{G}-\mathbf{G_A})\mathbf{C} - \mathbf{K}\mathbf{H} \end{bmatrix} \qquad (13$$

The effect of parametric uncertainty on stability robustness is computed by Monte Carlo evaluation of the eigenvalues of Eq. 10, with $\mathbf{F(p)}$, $\mathbf{G(p)}$, and $\mathbf{H(p)}$ substituted for $\mathbf{F_A}$, $\mathbf{G_A}$, and $\mathbf{H_A}$. Closed-loop eigenvalue densities portrayed on the stochastic root locus show the possible interaction of dynamic and estimator states, and the possible robustness degradation due to the estimator. Well-known loss of LQ stability margins when a state estimator is added [9] can be quantified by the probability of instability.

STOCHASTIC PERFORMANCE ROBUSTNESS

While stability is an important element of robustness analysis, performance robustness analysis is vital to determining whether important design specifications are met. Stochastic stability robustness is described by a single parameter, the probability of instability. Adequate performance -- initial condition response, response to commanded inputs, control authority, and rejection of disturbances is difficult to describe by a single scalar. However, elements of stochastic stability robustness analysis (e.g., Monte Carlo evaluation and use of the tests described above for confidence intervals) apply independent of the performance criteria chosen. This will be demonstrated in the sequel.

Numerous criteria stemming from classical control concepts exist as measures of adequate performance. Appealing to these, one can begin a smooth transition from stability robustness analysis to performance robustness analysis simply by analyzing the *degree of stability* or *instability* rather than strict stability. As described in [3], one method of doing this is to shift the vertical discriminant line from zero to Σ less than (or greater than) zero (Fig. 1a). Histograms and cumulative distributions for varying degrees of stability are readily given by the Monte Carlo estimate of $1 - \int_{-\infty}^{\Sigma} pr(\sigma)d\sigma$, where Σ represents a maximum real eigenvalue component, and $-\infty < \Sigma < \infty$. The histogram is a plot of $\frac{N[(\Sigma - \Delta) < \sigma_{max} \le \Sigma]}{J}$ vs. Σ; Δ is an increment in Σ, $N[.]$ is the number of cases whose maximum real eigenvalue components lie in the increment, and J is the total number of evaluations. The histogram estimates the *stability probability density function*, $pr(\Sigma)$, which is obtained in the limit for a continuous distribution of Σ as $\Delta \to 0$ and $J \to \infty$. The *cumulative probability distribution of stability*, $\Pr(\Sigma)$, is similarly estimated and presented as $\frac{N(\sigma_{max} < \Sigma)}{J}$ vs. Σ, the exact distribution being achieved in the limit as $J \to \infty$. Binomial confidence intervals are applicable to each point of the cumulative distribution, as there are just two values of interest, e.g., "satisfactory" or "unsatisfactory". The probability of instability is a special case where Σ equals zero.

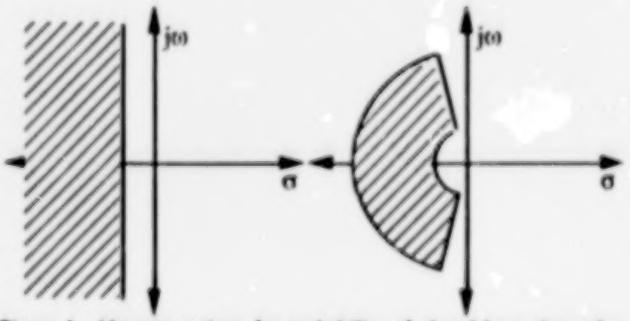


Figure 1 Alternate regions for probability of closed-loop eigenvalue location to aid in robustness analysis of transient response.
a) Alternate discriminant
b) Sector defined along lines of constant damping and natural frequency.

The measure of robustness resulting from the cumulative probability distribution is directly related to classical concepts

of rates of decay (growth) of the closed-loop response, time-to-half and time-to double:

$$t_{half} = \frac{0.693}{\zeta\omega_n} \quad (14$$

$$t_{double} = -\frac{0.693}{\zeta\omega_n} \quad (15$$

Taking degree-of-stability analysis one step further, rather than a vertical discriminant line, one can confine the closed-loop roots to *sectors in the complex plane* bounded by lines of constant damping and lines of constant natural frequency (Fig. 1b). Roots confined to these sectors would be expected to display a certain transient response speed. Again, the probability of roots lying within a sector is a binomial variable, and confidence interval calculations presented above apply.

While the speed of the transient response depends on the closed-loop poles, its magnitude and overall shape depends on the coefficients of the characteristic exponential and sinusoidal terms. Closed-loop zeros, residues, eigenvectors, and steady-state response are all concepts related to the magnitude of the response. The distribution of closed-loop zeros, of residues, of steady-state response, and of important elements of the eigenvectors all can be estimated by Monte Carlo analysis. Hence, a qualitative idea of the possible closed-loop responses can be obtained without calculating actual time responses.

Time responses provide the most clear-cut means of evaluating performance, but they are the most computation-intensive means as well. If actual time-responses are computed, stochastic performance robustness can be portrayed as a distribution of possible trajectories around a nominal or desired trajectory. After defining "envelopes" around the nominal trajectory (Fig. 2), the probability of violating the envelopes can be computed using Monte Carlo evaluation. The envelope chosen around the nominal trajectory encompasses scalar performance measures; the trajectories in Fig. 2 are examples of bounds defined by minimum and/or maximum allowable dead time, delay time, rise time, time to peak overshoot, peak overshoot, settling time, and steady-state error. While it is simple to conclude that a response violates such an envelope, individual responses within the envelope may not be acceptable. In such cases, the derivative of a response and envelopes around the derivative also can be used as performance criteria.

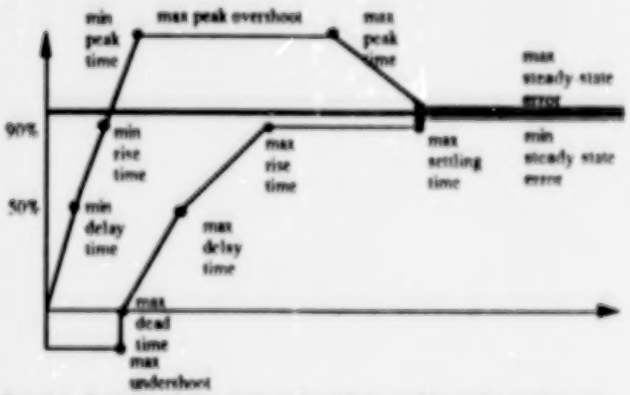


Figure 2. Example of step response bounds formed by scalar performance characteristics.

There is no unique set of criteria defining envelopes that bound an acceptable time response; the segmented envelopes in Fig. 2 can be smoothed, or other scalars can be used to define points on the envelope. However, once an envelope is defined, time response distributions can be computed by Monte Carlo methods. The closed-loop time response to a command input, disturbance, initial condition, or some combination is evaluated J times, and for each evaluation, the trajectory is a binomial variable; it either stays within the envelope or violates the envelope. Although computing actual time responses is more computation-intensive than probability-of-instability estimation or estimation of scalar values associated with performance robustness, such analysis is well within the capability of existing workstations.

STOCHASTIC ROBUSTNESS AS A CONTROL DESIGN AID

Stochastic robustness metrics can and should be related to control system design parameters for robust control system design. While general "rules of thumb" regarding the design of robust control systems are useful, stochastic robustness metrics can identify non-obvious robustness "structures" of particular applications. Figure 3 illustrates tradeoffs that can exist and be uncovered by stochastic robustness analysis. Consider Fig. 3a, which shows the upper-half plane of a plant that has a complex pair of poles and a right-half-plane zero. Hypothetical "uncertainty circles" are drawn around possible closed-loop-root locations. As gain increases along the root locus, the uncertainty is magnified, and stability robustness decreases, with possible closed-loop root locations in the right-half plane at high enough gain. This case illustrates one where the decrease in robustness may be monotonic. Figure 3b postulates a system with a real pole and a complex pair of poles and zeros. In Fig. 3b, the hypothetical root locus nears the $j\omega$ axis before ending at the zero in the left-half plane. Again, uncertainty circles enlarge as gain increases. In this case, it is possible that the root distributions cross into the right-half plane, yet are entirely in the left-half plane as gain increases further. Here, stability robustness (as measured by the probability of instability) may have local or global minima (as a function of gain). For multivariable systems with many parameters, the intrinsic structure of the problem and the tradeoff between the spread in closed-loop-root-location uncertainty versus the magnitude of the control gains may not be immediately evident. Plots of stochastic robustness metrics versus scalar control design parameters provide the necessary insight.

Figure 3 Illustration of design insight revealed by stochastic robustness analysis. Solid points indicate closed-loop eigenvalues enclosed by "uncertainty circles".
a) Root locus sketch where stability robustness decreases monotonically with increased gain.
b) Root locus sketch where stability robustness decreases then increases with increased gain.

EXAMPLE: SINGLE LINK ROBOT ARM

As an example of the concepts discussed above, stochastic robustness is applied to the flexible one-link robot described in Ref. 10. The model retains the first three flexible modes, and the tip of the link is controlled by applying a control torque to the hub, or base of the link. Because the structure of this model is representative of a general flexible structure, and the physical parameters are easily identifiable, it is a good candidate for stochastic robustness analysis.

The dynamic, control effect, and output matrices are given by

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \omega_1^2 & -2\zeta_1\omega_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \omega_2^2 & -2\zeta_2\omega_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \omega_3^2 & -2\zeta_3\omega_3 \end{bmatrix} \quad (16$$

$$G = \frac{1}{I_T}\begin{bmatrix} 0 \\ 1 \\ 0 \\ \phi_1'(0) \\ 0 \\ \phi_2'(0) \\ 0 \\ \phi_3'(0) \end{bmatrix} \quad H = \begin{bmatrix} L & 0 & \phi_1(L) & 0 & \phi_2(L) & 0 & \phi_3(L) & 0 \\ 0 & 1 & 0 & \phi_1'(0) & 0 & \phi_2'(0) & 0 & \phi_3'(0) \end{bmatrix}$$

(17,18

where x is the length along the arm, $\phi_i(x)$,are the normal modes, $\phi_i' = \frac{d\phi_i}{dx}$, L is the length of the arm, and I_T is the total inertia of the arm. The measurements taken through **H** are the tip displacement and hub-rate, respectively. The flexibility of the open-loop system is apparent in the open-loop eigenvalues, which are 0, 0, -0.177 ± 11.81j, -0.432 ± 21.61j, and -0.968 ± 48.37j. The system has a readily identifiable 14-element parameter vector:

$$p = [\zeta_1\ \omega_1\ \zeta_2\ \omega_2\ \zeta_3\ \omega_3\ \phi_1'(0)\ \phi_2'(0)\ \phi_3'(0)\ L\ \phi_1(L)\ \phi_2(L)\ \phi_3(L)\ I_T\ ] \quad (19$$

Details concerning the modeling and parameter identification are given in [10]. The linear-quadratic-regulator designed in [10] was used for demonstration and review of stochastic stability robustness analysis. The quadratic performance index weights tip position and tip-rate, and the LQR state-weighting and control-weighting matrices are

$$Q = 0.01F^TH^T\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}HF + H^T\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}H,\ R = 0.001 \quad (20, 21$$

resulting in linear-quadratic regulator gains

$$C = [35.42\ \ 13.38\ \ 41.24\ \ 2.65\ \ 50.32\ \ -0.67\ \ 135.46\ \ 1.52] \quad (22$$

and closed-loop eigenvalues -5.41 ± 48.8j, -6.47 ± 23.8j, -6.1 ± 2.66 j, -7.7 ± 11.42j.

A uniform probability density function was chosen to model the parameter uncertainty statistics, with variations between ±2% of the nominal values for L and I_T and ±25% for the remaining parameters. The stochastic root locus for 50,000 evaluations of the full-state feedback case is given in Fig. 4. The nominal eigenvalues are marked, and the possible distribution is indicated by the height above the complex plane in units of roots/length along the real axis and roots/area in the complex plane. The "bin" size in Fig. 4 is 0.9 along the axis and 0.9 x 0.9 off of the axis. The probability of instability estimate ($\hat{P}$) for 50,000 evaluations is zero. Each of the four complex eigenvalue pairs appears in Fig. 4 as a "peak", with some distribution around the peak due to parametric uncertainty. The peaks can be well-defined (as in the lowest frequency complex pair) or broad (as in the highest frequency pair), and the nominal eigenvalues are not necessarily at the peaks of the distributions. Uncertainties of the magnitude chosen cause complex pairs to coalesce into real roots resulting in a distribution along the real axis.

Degree of stability is portrayed in Fig. 5 by the histogram and cumulative distribution of stability around the origin for a bin size of 0.25. (The smaller bin size was chosen for better resolution.) For binary parameter variations [3] of the same magnitude as the maximum uniform variations, 2^{14} or 16,384 deterministic evaluations give a zero probability of instability as well. These results indicate good stability robustness in the face of reasonably large uncertainties.

Moving to performance robustness analysis, Fig. 6 shows a top, contour-shaded view of a stochastic root-locus with sector bounds defined by $4 \le \omega_n \le 65$ and $\zeta > 0.1$. For 50,000 evaluations, the probability of having closed-loop eigenvalues outside of these bounds is 0.0147, with 95% confidence intervals (L,U) = (0.0136,0.0158). While the shape of the time response depends on closed-loop zeros, a minimum speed of response can be guaranteed by requiring that all closed-loop eigenvalues lie within a specified sector.

Figure 7a presents example step response envelopes and the response of the tip to a 4.8 cm position command input for 500 Monte Carlo evaluations. The control time history corresponding to the mean response is given in Fig. 7b. The transfer function between hub torque and tip position is non-minimum phase; thus the step response exhibits an initial response in the wrong direction. The time response envelopes in Fig. 7 indicate the maximum acceptable non-minimum phase response. For 500 responses, the probability of violating the time response envelope is 0.184 with 95% confidence intervals (L,U) = (0.151,0.221). Individual responses characteristic of those evaluated by Monte Carlo analysis are given in Fig. 8. While the responses fill out the envelope, some of the individual responses within the envelope may not be acceptable in the face of real-world criteria governing rate of change of the response (Fig. 8c). This is a case where checking envelopes around the derivative of the response may be necessary. Similar analyses can be performed on control trajectories to make sure bandwidth and control effort limitations are not violated during the simulation.

It is instructive from a design standpoint to plot stochastic robustness measures versus design parameters used to calculate feedback gains. Since there is only a single control in this example, the control weighting matrix R is a scalar and can be used as the design parameter. Two stochastic performance robustness measures are plotted versus R in Fig. 9 - the probability of violating the time response envelope and the probability of degree of instability. As control gains increase, the closed-loop roots are pushed farther into the left half plane, but they tend to migrate farther from their nominal values as well. At some value of control gain, there is a tradeoff between how far roots migrate and their location in the left half plane; thus a local minimum is apparent around R = 0.001 in the probability-of-degree-of-instability curves. While degree of instability improves for very small R, the control gains for this case are unrealistically large. For larger R (smaller control gains), the nominal closed-loop roots have real parts in the range of the values of Σ used; thus the probability of degree of instability increases rapidly beyond R = 0.01. The probability of the violating time response envelopes takes on the same shape as a function of R as the probability of degree of instability. In this example, the probability of instability is zero for all values of R checked, however, minima in the probability of instability versus design parameter curve can occur as well [11].

Figure 10 shows the stochastic root locus for the LQG system with estimator gains based on disturbance effect matrix L = G and disturbance and noise covariance matrices

$$W = 1,\ N = \begin{bmatrix} 0.005 & 0 \\ 0 & 10 \end{bmatrix} \quad (23, 24$$

With estimator states added, the stochastic root locus changes in overall character from the full-state feedback case. Peaks are sharper, and the distribution along the real axis is less pronounced. In particular, note the eigenvalues associated with the largest peaks. In the full-state feedback system (Fig. 4) a broad distribution is associated with these eigenvalues, yet in Fig. 10, this pair of eigenvalues undergo very little variation from their nominal values! While the extent of the distribution into the left half plane is about the same as in Fig. 4, LQG system eigenvalues do migrate into the right half plane. The probability of instability estimate and 95% confidence intervals for 50,000 evaluations are $\hat{P}$ = 0.0771, and (L,U) = 0.0748, 0.0795, representing a significant loss in the stability robustness characteristic of the LQR system. These results reflect the well known fact that guaranteed stability margins of an LQR system are lost when an estimator is added.

Loop Transfer Recovery (LQG/LTR) [12] is a common design technique by which the trade-off between estimator performance and stability robustness can be made systematically using a single design parameter. Figure 11 illustrates the LTR mechanism for this example using the probability of instability as the robustness measure. Here plotting the probability of instability as a function of design parameter v ($W = vW_0$) shows that there is a value of v (v = 2) that minimizes the probability of instability. The fact that such a minimum exists may not be apparent by simply examining the estimator eigenvalues. The kind of results presented in Fig. 11 offer design insight and show robustness characteristics not revealed by other robustness metrics.

CONCLUSIONS

Stochastic robustness offers a rigorous yet straightforward alternative to current metrics for control system robustness that is simple to compute and is unfettered by normally difficult problem statements, such as non-Gaussian statistics, products of parameter variations, and structured uncertainty. Principles behind stochastic robustness can be applied to scalar performance metrics as well as time responses, making it a good candidate for overall robustness analysis. Both performance and stability metrics resulting from stochastic robustness analysis can provide details relating robustness characteristics intrinsic to a given application and scalar control design parameters, making it a good candidate for optimization techniques as well.

ACKNOWLEDGMENTS

This research has been sponsored by the FAA and the NASA Langley Research Center under Grant No. NGL 31-001-252 and by the Army Research Office under Grant No. DAAL03-89-K-0092.

Figure 4 Stochastic root locus for the single-link robot with uniform parameters, 50,000 evaluations.

REFERENCES

1) Stengel, R.F., "Some Effects of Parameter Variations on the Lateral-Directional Stability of Aircraft," AIAA *Journal of Guidance and Control*, Vol. 3, No. 2, pp. 124-131, Apr 1980.

2) Stengel, R.F., and Ryan, L.E., "Stochastic Robustness of Linear Control Systems", *Proceedings of the 1989 Conference on Information Sciences and Systems*, pp. 556-561, March, 1989.

3) Stengel, R.F., and Ryan, L.E., "Multivariate Histograms for Analysis of Linear Control System Robustness", *American Control Conference Proceedings*, pp. 937-943, Pittsburgh, PA, June 1989.

4) Ryan, L.E., and Stengel, R.F., "Application of Stochastic Robustness to Aircraft Control Systems", *AIAA Guidance, Navigation and Control Conference Proceedings*, pp. 698-708, Boston, Mass, Aug. 1989.

5) Conover, W.J., *Practical Non-parametric Statistics*, John Wiley and Sons, New York, 1980.

6) Anderson, T.W., and Burnstein, H., "Approximating the Upper Binomial Confidence Limit", *Journal of the American Statistical Association*, Vol. 62, pp. 857-861, Sept 1967.

7) Anderson, T.W., and Burnstein, H., "Approximating the Lower Binomial Confidence Limit", *Journal of the American Statistical Association*, Vol. 63, pp. 1413-1415, Dec 1968.

8) Stengel, R.F. *Stochastic Optimal Control: Theory and Application*, John Wiley and Sons, New York, 1986.

9) Doyle, J.C., "Guaranteed Margins for LQG Regulators," IEEE *Transactions on Automatic Control*, Vol. AC-23, No. 4, pp. 756-757, Aug 1978.

10) Cannon, R.J., and Schmitz, E., "Initial Experiments on the End-Point Control of a Flexible One-Link Robot", *The International Journal of Robotics Research*, Vol. 3, No. 3, pp. 62-75, Fall, 1984

11) Stengel, R.F., and Ryan, L.E., "Stochastic Robustness of Linear Time-Invariant Systems", to appear in *IEEE Transactions on Automatic Control*.

12) Doyle, J.C., and Stein, G., "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis," IEEE *Transactions on Automatic Control*, Vol. AC-26, No. 1, pp. 4-16, Feb 1981.

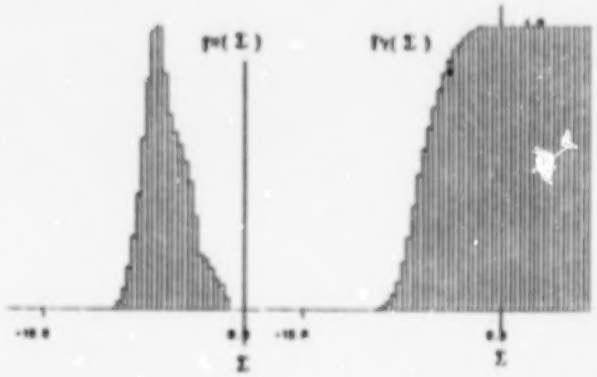


Figure 5 Histogram and cumulative distribution for the single-link robot with uniform parameters, 50,000 evaluations.

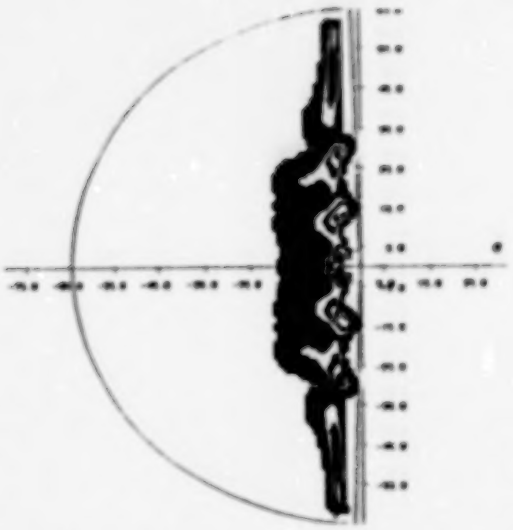


Figure 6 Top view of a stochastic root locus with sector bounds for the single-link robot, 50,000 evaluations.

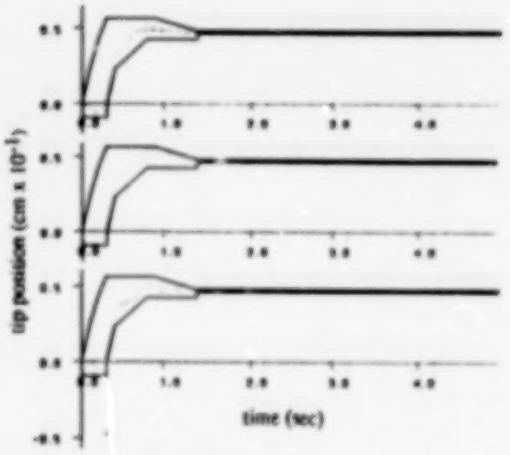


Figure 8 Examples of individual tip responses
a) Acceptable response within envelope.
b) Response violates envelope.
c) Response is within envelope, but criteria governing its derivative may be required.

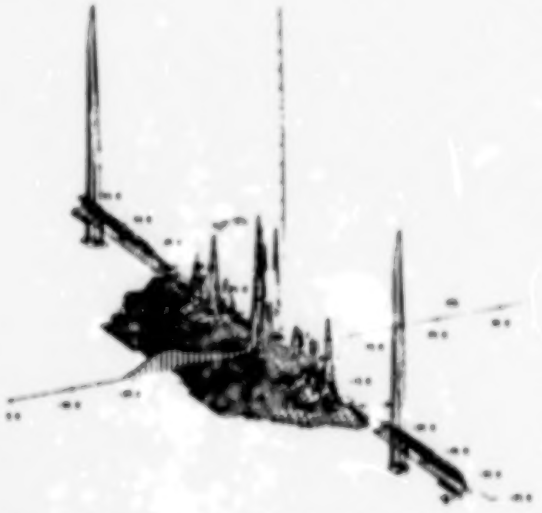


Figure 10 Stochastic root locus for the single-link robot with state estimation (LQG), 50,000 evaluations.

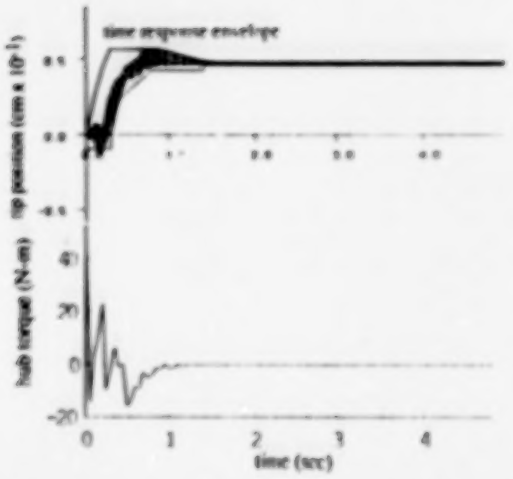


Figure 7 Time histories associated with tip position command of 4.8 cm.
a) 500 Monte Carlo evaluations of the tip response. Envelopes are defined by scalar performance criteria. Nominal response is indicated by the solid line.
b) Nominal control input.

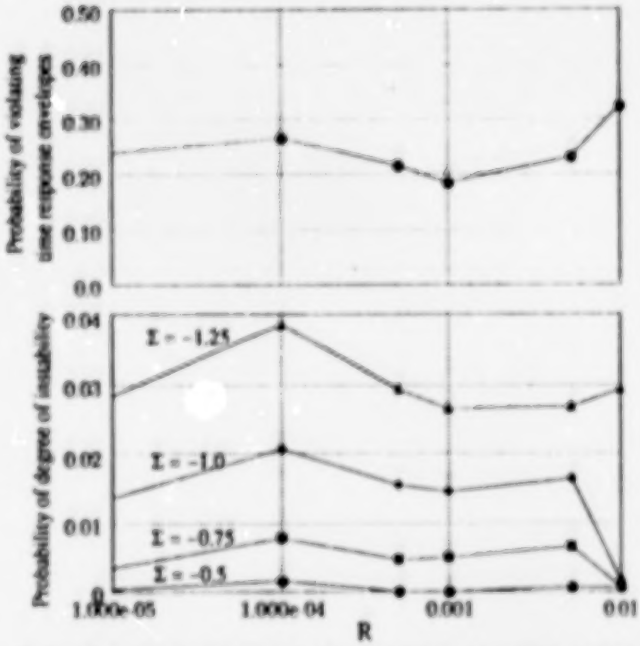


Figure 9 Stochastic performance robustness metrics vs. control weighting matrix R.
a) Probability of violating time response envelopes.
b) Probability of degree of instability, for values Σ along the real axis.

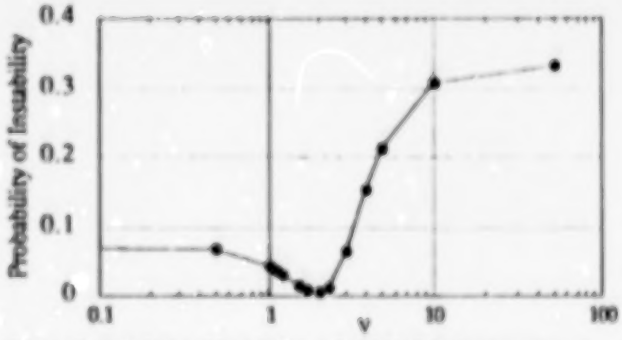


Figure 11 Probability of instability vs. LQG/LTR design parameter v.

SYSTEM IDENTIFICATION FOR NONLINEAR CONTROL USING NEURAL NETWORKS

Robert F. Stengel[1] and Dennis J. Linse[2]
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, NJ

ABSTRACT

An approach to incorporating artificial neural networks in nonlinear, adaptive control systems is described. The controller contains three principal elements: a nonlinear-inverse-dynamic control law whose coefficients depend on a comprehensive model of the plant, a neural network that models system dynamics, and a state estimator whose outputs drive the control law and train the neural network. Attention is focused on the system-identification task, which combines an extended Kalman filter with generalized-spline function approximation (e.g., basis splines, back-propagation feedforward networks, or cerebellar model articulation controller networks). Continual learning is possible during normal operation (without taking the system off line for specialized training). Nonlinear-inverse-dynamic control requires smooth derivatives as well as function estimates, imposing stringent goals on the approximating technique.

INTRODUCTION

Current research on artificial neural networks has been spurred by the need for pattern recognition, optimization, and control algorithms that can learn from experience. The biological paradigm provides an attractive starting point, as learning is the crux of living activity. Nevertheless, it remains to be seen whether or not biologically-motivated algorithms (i.e., artificial neural networks) are computationally appealing, particularly as compared to more conventional numerical techniques for function approximation. These techniques also can "learn" or "be taught," although those terms rarely are used for the process of fitting curves or determining coefficients from empirical data.

The most important question appears to be how well neural networks generalize multivariate functions of high input dimension. Basis splines (or B-splines) represent an efficient, modern, conventional technique [1, 2]. A univariate B-spline function approximation can be trained quickly and accurately using deterministic or averaging techniques, and the resulting spline approximant has nice numerical properties, such as positivity and compact support. Multivariate B-splines can be computed, but they may become cumbersome in both training and operation if the number of independent variables is much larger than two. Furthermore, the multivariate B-spline requires a fairly rigid definition of the input space, resulting in potentially high computation times and storage requirements. Neural networks can, in principle, gain efficiency through automatically identifying and concentrating on significant receptive regions in the input space [3, 4].

Neural networks can be incorporated in nonlinear control logic in at least two ways. The first is as a *generalized feedforward control function* that maps the space of desired command responses into corresponding control settings. This neural network is conceptually straightforward, as it is based on the inverse of the plant's response to control inputs [5]. The second is to use the neural network as a *model of plant dynamics*, which is embodied in the control law but is distinct from the command/control specification [as in nonlinear-inverse-dynamic (NID) control]. The first approach may lead to more compact implementation, while the second may provide more flexibility for multi-modal or fault-tolerant control.

In the remainder of the paper, we discuss elements of the second approach. After briefly stating the nonlinear control problem and its NID solution, a learning structure for neural networks based on an extended Kalman filter is described, and details of two network formulations are presented. It is shown that smooth neural network outputs are desirable and, in fact, required for NID control. The back-propagation feedforward network has sufficient smoothness as normally defined, but the cerebellar model articulation controller network must be modified to eliminate quantization effects.

THE NONLINEAR CONTROL PROBLEM

Consider the control of a nonlinear dynamic plant as described by,

$$\dot{x} = f(x) + g(x, u) + w \tag{1}$$

$$y = h(x) \tag{2}$$

$$z = h_z(x) + n \tag{3}$$

with x and $w \in R^n$, u, $y \in R^m$, and z and $w \in R^l$. f and g are smooth vector fields in R^n; h and h_z are smooth vectors fields in R^m and R^l. x represents the dynamic state, u is its control, w is the disturbance input, y is the response vector, z is the measurement of the process, and n is the measurement error.

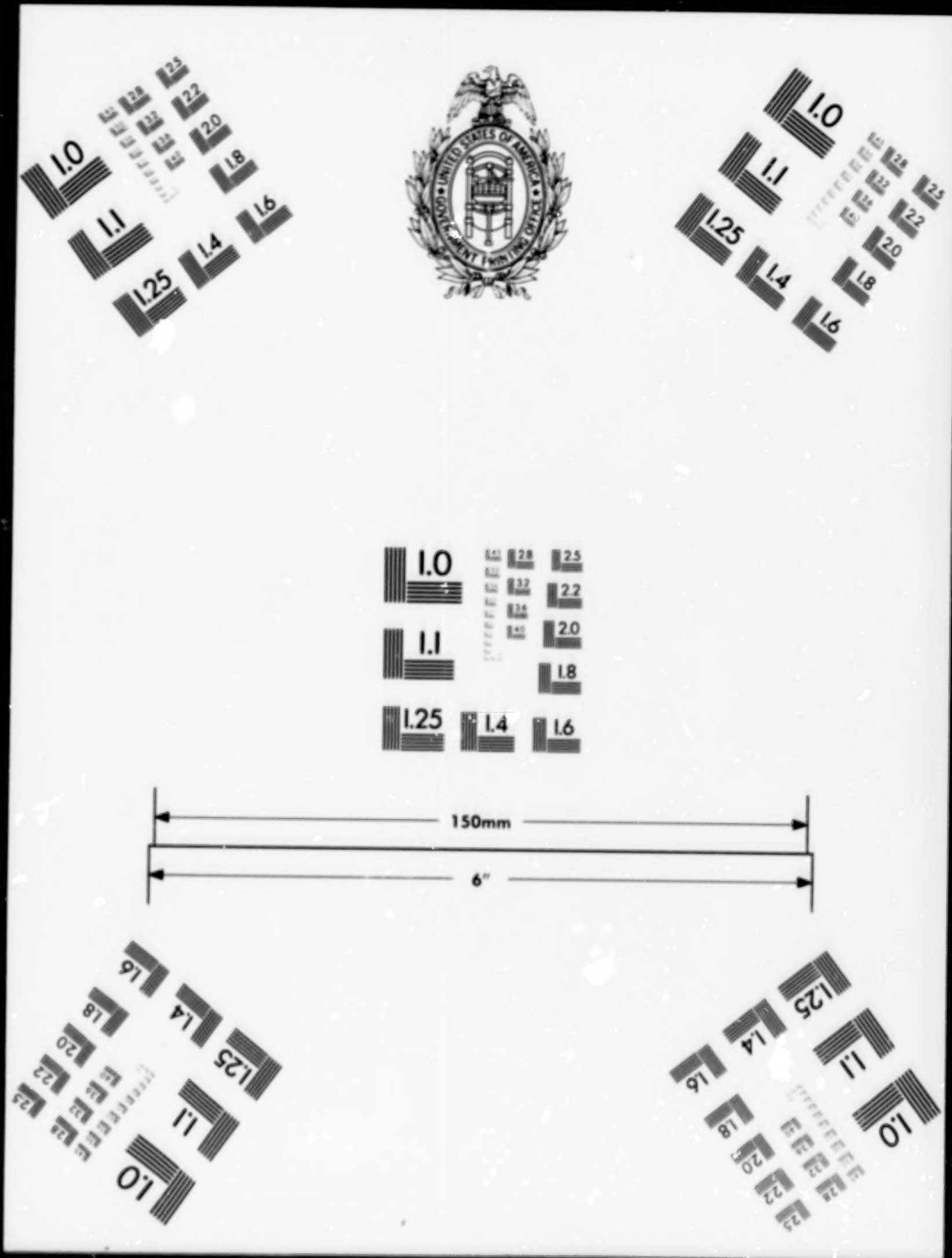
For illustration, assume that eq. 1 and 2 can be written as,

$$\dot{x} = f(x) + G(x)\, u + w \tag{4}$$

---

[1]Professor
[2]Graduate Research Assistant

Presented at the 1990 Conference on Information Sciences and Systems, Princeton, NJ, March, 1990.


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
150mm
6"

$$y = H x \tag{5}$$

with $w = 0$ and $z = x$. Define the *derivative response vector* $y^{(d)}$ by termwise differentiation of eq. 5 and substitution of eq. 4, repeated until the control affects each element of $y^{(d)}$ linearly [6, 7]:

$$y^{(d)} = f^*(x) + G^*(x)\, u \tag{6}$$

$G^*(x)$ is guaranteed to be nonsingular, so it is possible to construct an *integrator-decoupling control law* of the form,

$$u = c(x) + C(x)v \tag{7}$$

where,

$$C(x) = [G^*(x)]^{-1} \tag{8}$$

$$c(x) = -[G^*(x)]^{-1} f^*(x) \tag{9}$$

and v is a command input vector. Together with additional compensation to provide satisfactory closed-loop stability and command response, eq. 7 forms a *nonlinear-inverse-dynamic control law*.

For typical response vector specifications in aircraft control applications [8, 9], one to three differentiations are required to bring out the control effect in $y^{(d)}$; hence, up to two partial derivatives of $f(x)$ and $G(x)$ (with respect to x) must exist in those cases. Our goal is to approximate an uncertain but observable portion of the system model using neural networks. Consequently, neural network outputs must be smooth enough for the differentiation and inversion contained in eqs. 6, 8, and 9 to take place.

THE SYSTEM IDENTIFICATION PROBLEM

Consider a system described by eq. 1-3 with $w = 0$ and $z = x$, in which $f(x)$ is known without error and $g(x,u)$ is subject to uncertainty or change. Let $g(t)$ represent the value of $g(x,u)$ at time t during some period of system operation; then $g(t)$ could be estimated precisely as

$$g(t) = \dot{x}(t) - f[x(t)] \tag{10}$$

and $[g(t), x(t), u(t)]$ would form a suitable training set for a functional approximation to $g(x,u)$. During the course of normal system operation, it is assumed that $x(t)$ and $u(t)$ would provide a succession of training sets that were sufficiently rich for function approximation, and it would be possible to "replay" system trajectories off line for improved learning. There is no restriction on the form of $g(x,u)$; it is suggested here that $g(x,u)$ be represented by one or more *neural networks*.

More realistically, w and n are not zero, and z is not x, so eq. 10 cannot be evaluated. However, with the usual observability requirements, it is possible to make estimates of $g(t)$ and $x(t)$ using an *Extended Kalman Filter*, as described, for example, in [10]. Following a procedure analogous to *Estimation Before Modeling* [11-15], the state vector is augmented to include $g(t)$ and enough derivatives of $g(t)$ to assure a smooth estimate of $g(t)$. [In the off-line aircraft system identification applications presented in [11-15] (Fig. 1), $g(t)$ was treated as an integrated random-walk (Gauss-Markov) process, necessitating the estimation of $\dot{g}(t)$ and $\ddot{g}(t)$ as well as $g(t)$. Improved estimates were obtained by following the filtering step with a reverse-time modified Bryson-Frazier (MBF) smoothing step [16].] In the current application, the filtered estimates $\hat{g}(t)$ and $\hat{x}(t)$ are used with $u(t)$ to train the neural network $\hat{g}(\hat{x}, u)$. Smoothed estimates could be used for further off-line training improvement. [See [17] for a related application of filtering and smoothing].

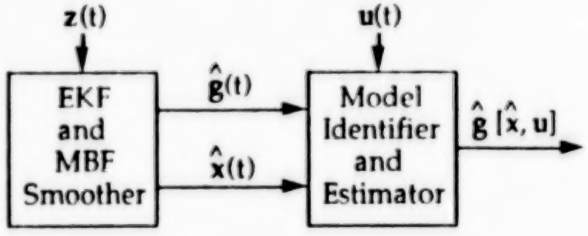


Figure 1. Basic Estimation Before Modeling.

For the airplane example, $x(t)$ is an 11-component vector representing 3 components (each) of translational and rotational rates, 4 quaternion components, and altitude. $f[x(t)]$ represents kinematic and inertial effects, while the vector $g(t)$ has just 6 nonzero terms due to aerodynamic, thrust, and control effects. These terms are axial, side, and normal specific forces [$X(t)$, $Y(t)$, and $Z(t)$] and roll, pitch, and yaw specific moments [$L(t)$, $M(t)$, and $N(t)$]. Consequently, 6 parallel neural networks would be formed for this application. Aerodynamic effects can be normalized using air density, airspeed, reference area, and reference length; hence, the neural networks would represent the non-dimensional coefficients $C_X(x,u)$, $C_Y(x,u)$, $C_Z(x,u)$, $C_l(x,u)$, $C_m(x,u)$, and $C_n(x,u)$.

The neural network can be integrated with NID control, as shown in Fig. 2. The dashed arrow from the estimation model block to the nonlinear control law block represents an update of the functional form of the control law rather than the direct feedback path of the other connections.

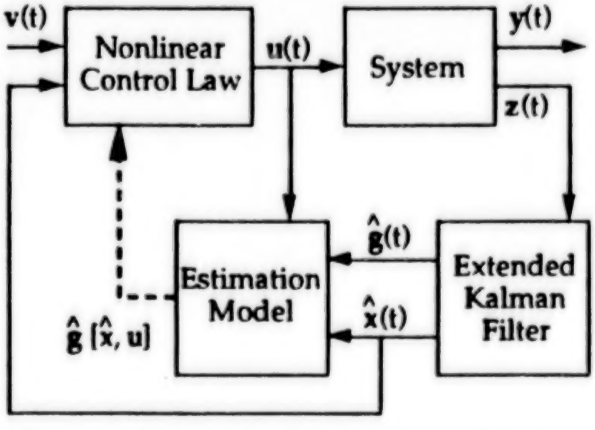


Figure 2. Integration of Neural-Network System Identification with Nonlinear-Inverse-Dynamic Control.

NEURAL NETWORK ESTIMATION SCHEMES

Back-propagation feedforward neural networks and CMAC neural networks are particularly suitable for system identification. Each represents a different structural model of a process and presents unique capabilities for the system identification process. Although neural network techniques appear very different from classic techniques such as B-splines, under close examination there are strong similarities. As discussed in [18], most neural networks, including the two models discussed here, can be viewed as generalized splines. This observation opens the door on a large body of knowledge about function approximation. A brief description of the two networks follows. Further details about each can be found in [19].

Back-Propagation Feedforward Network

One of the simplest, and most used, artificial neural network is the back-propagation feedforward network [20]. A feedforward neural network is constructed from a weighted interconnection of simple nonlinear elements called *nodes*. The nodes are separated into *layers*, with the input to each node exclusively from the previous layer and the output connected only to the following layer. The orderly construction leads to a simple recursive definition

$$\mathbf{t}^{(k)} = \mathbf{W}^{(k-1)}\,\mathbf{x}^{(k-1)} \tag{11}$$

$$\mathbf{x}^{(k)} = \mathbf{s}^{(k)}\left[\mathbf{t}^{(k)}\right] \tag{12}$$

for layers k = 1 to N. The *activation function* vector, $\mathbf{s}^{(k)}[\mathbf{t}]$, is formed by concatenating the scalar *activation functions* contained in each node of a single layer:

$$\mathbf{s}^{(k)}[\mathbf{t}] = [\sigma_1(t_1) \ldots \sigma_n(t_n)]^T \tag{13}$$

One input to each layer is fixed at unity and serves as a threshold, as illustrated by Fig. 3.

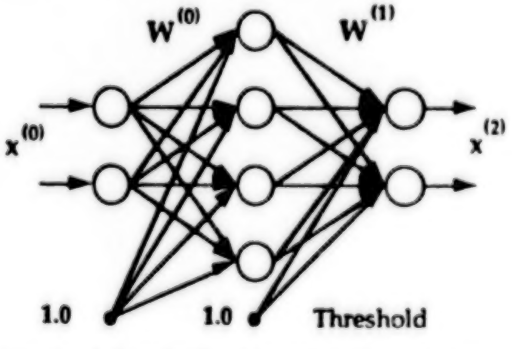


Figure 3. A Simple Feedforward Neural Network.

Given the form of the activation function (usually the same in each node), the number of layers, and the number of nodes in each layer, the *back-propagation rule* [20] can be used to update the values of the weights, $\mathbf{W}^{(k)}$. The back-propagation rule attempts to minimize the mean-square error of the difference between the network output and a desired output value using a gradient-based descent algorithm. The choice of activation functions and numbers of layers and nodes is the subject of much research [e.g. [21]].

Cerebellar Model Articulation Controller

The Cerebellar Model Articulation Controller (CMAC) neural network [3, 4] is a generalization of a simple table look-up of a multi-input/single-output nonlinear function. Originally modeled on the uniform structure of the cerebellum, the basic CMAC divides the multi-dimensional input space into a series of overlapping hypercubes. A weight is assigned to each hypercube. The output of the CMAC for a specific input point is simply the sum of the weights of those hypercubes enclosing the input point. Applications to date have been in the areas of function approximation [3, 4, 17], robotics [5,22,23], and adaptive control [24].

Introducing an *association space*, $\mathcal{A}$, the operation of the CMAC is best described by two mappings. The first goes from the input space, $\mathcal{X}$, to the association space

$$Q: \ \mathcal{X} \rightarrow \mathcal{A} = \{0, 1\}^{N_A} \tag{14}$$

The elements of the binary *selector vector*, **a**, in the N_A dimensional $\mathcal{A}$ are fixed by

$$a_j = \begin{cases} 1 & x \in \Lambda_j \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where Λ_j is the j^{th} receptive region. The *receptive regions* are the overlapping hypercubes that cover the input space. By overlapping the receptive regions such that any input lies within exactly C receptive regions, **a** has C non-zero values. The size of C controls the generalization capability of the CMAC.

The second mapping from **a** to the scalar output, y,

$$\mathcal{R}: \ \mathcal{A} \rightarrow \mathcal{Y} \tag{16}$$

is simply a summation of the weights associated with the selected receptive regions. Defining a weight vector, **w**, of dimension N_A, the inner product of **w** with **a** generates the desired output

$$y = \mathbf{w}^T\mathbf{a} \tag{17}$$

The basic CMAC (Fig. 4) quantizes the inputs based on the size of the overlap of the receptive regions. Since there is only one further linear operation performed, the output of the CMAC is a piecewise-continuous function.

Since the weights and output are connected by a simple linear operation, a learning algorithm is easy to prescribe. Each weight contributing to a particular output value is adjusted by a fraction of the difference between the network output and the desired output. The fraction added is determined by both the speed of desired learning and the number of weights contributing to the output.

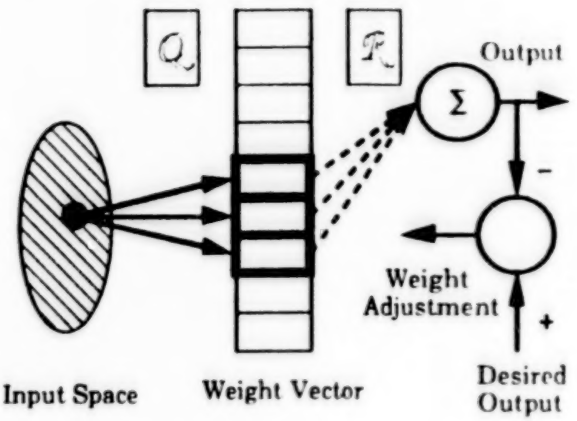


Figure 4. The Basic CMAC Architecture.

By its construction, the fidelity of the learned CMAC response is directly related to the quantization of the input space into receptive regions. For multi-dimensional input spaces, the number of receptive regions, and therefore the number of weights, scales as

$$N_A = \vartheta(Q^n) \tag{18}$$

where Q is the number of number of divisions in each input dimension and n is the number of input dimensions. For even reasonably sized, multi-input problems, the number of weights quickly overwhelms the memory capacity of most computers. Fortunately, in many applications, such as control, only small regions of (or trajectories through) the input space are used. By implementing a many-to-few mapping of the weight vector using a random hash-coding technique [25], the number of necessary weights can be compressed to a reasonable size. For example, in [23], a CMAC was applied to a robotics problem with 15 inputs. The weight vector was compressed using a random hash function from the original large length down to lengths between 32,768 and 8 weights. Virtually no degradation in performance was visible for lengths down to 1024.

DIFFERENTIABILITY OF THE ESTIMATION SCHEMES

The application to NID control imposes strong restrictions on the estimation scheme to be used, including the the existence of smooth derivatives. The two neural network techniques described have differing capabilities in this regard.

Recognizing that the overall function of the feedforward network is

$$\mathbf{x}^{(N)} = \mathbf{g}\left[\mathbf{x}^{(0)}\right] \tag{19}$$

the derivative of the network function, $\partial \mathbf{g}/\partial \mathbf{x}^{(0)}$, can be calculated recursively as

$$\frac{\partial \mathbf{x}^{(k)}}{\partial \mathbf{x}^{(k-1)}} = \frac{\partial \mathbf{s}^{(k)}[\mathbf{t}^{(k)}]}{\partial \mathbf{t}^{(k)}} \mathbf{W}^{(k-1)} \tag{20}$$

Since the weight matrices are constant, the differentiability of the network is determined by the activation function used in each node. One of the most common activation functions is the sigmoid

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad \forall\, z \in R \tag{21}$$

Since this function is infinitely differentiable, the full network also is infinitely differentiable.

The basic CMAC network is very different from the feedforward network. The quantization of the input space causes the CMAC to have piecewise-continuous output as demonstrated in Fig. 5 for a coarse quantization of the input value.

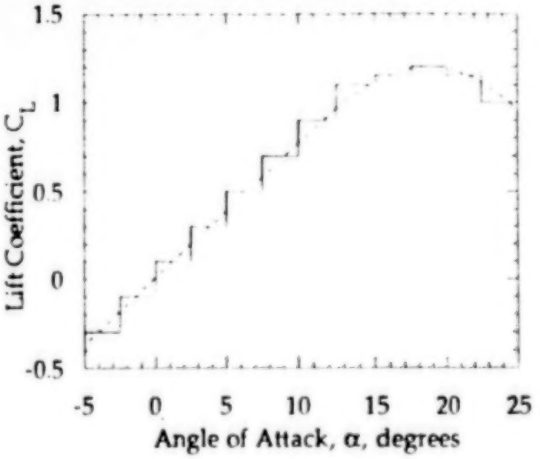


Figure 5. Example CMAC output for very coarse receptive regions.

Although useful in many other situations, the basic CMAC cannot be used in NID control because the control laws require continuous derivatives. Noticing the similarity between the receptive regions of CMACs and first-order basis splines, Lane et al. introduce *higher-order CMACs* in [26]. Equation 15 is redefined to be

$$a_j = \begin{cases} B_{n,j}(x) & x \in \Lambda_j \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

where the basis functions, B, of order, n, differ from conventional B-splines [basis functions] [1, 2, 26] to account for differing support (or receptive) regions. Due to the similarity of the basis functions, higher-order CMACs inherit the continuity properties of B-splines, including (n - 2) continuous derivatives. The penalty paid for this increase in continuity is the increased computation necessary in eq. 22 when compared to eq. 15. An example of receptive regions for a higher-order CMAC is given in Fig. 6.

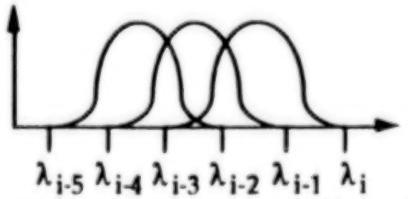


Figure 6. Receptive Regions of a Higher-Order CMAC.

EXAMPLE NEURAL NETWORK SYSTEM IDENTIFICATION

As a preliminary step to the full implementation of the integrated neural-network, NID control law, a simple investigation of the capabilities of the neural networks was performed. A one-dimensional test function representing lift coefficient variation with angle-of-attack for the airplane example is given in Fig. 7.

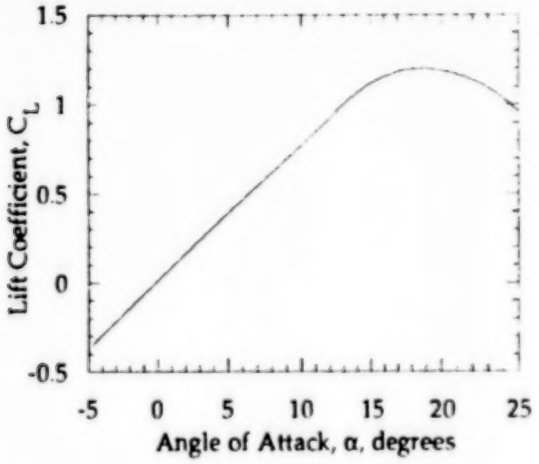


Figure 7. Lift curve test function

The learning capabilities of neural networks were tested by presenting random angles-of-attack between –5° and 25°, and the corresponding lift coefficients to each network. The result after 50,000 random training samples presented to a back-propagation network with two hidden layers of 10 and 5 nodes respectively, is given in Fig. 8. The error between the network output and the actual lift curve is very small, indicating a good representation of the original function.

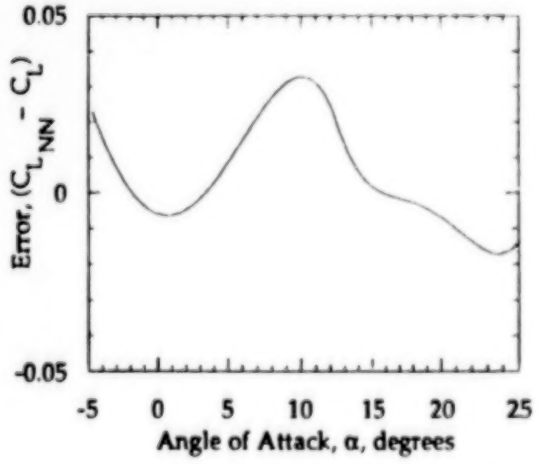


Figure 8. Back-propagation error for lift curve test function

The basic CMAC network network is similarly tested by presenting random input/output pairs. The result after 2000 such presentations for a network with 100 weights and a generalization value, C, of 10 (Fig. 9) also is very good. The noisy appearance of the network error is due to the piecewise continuity of the CMAC output. Higher-order CMACs will alleviate this problem.

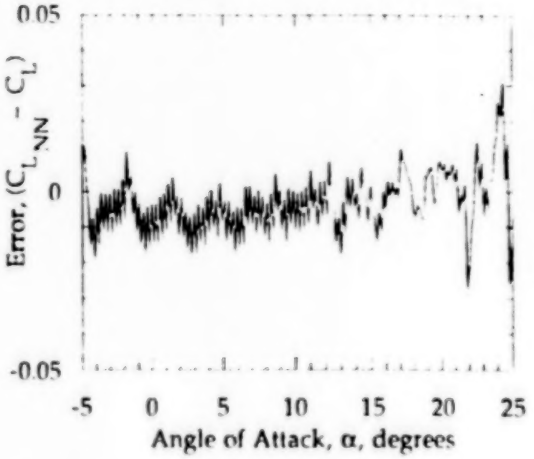


Figure 9. CMAC error for lift curve test function

Further testing of the neural networks, including multi-dimensional inputs and higher-order CMACs, remains to be completed. Full implementation of the integrated neural-network, NID control law (Fig. 2) will follow.

CONCLUSIONS

Neural networks provide a feasible alternative for function approximation in adaptive (or learning) nonlinear control systems. Parallel neural networks representing generalized forces due to state and control can be trained using the outputs of an extended Kalman filter. Neural network outputs must be continuously differentiable for use with nonlinear-inverse-dynamic control logic. Back-propagation feedforward neural networks using sigmoidal activation functions have the necessary differentiability. Cerebellar model articulation controller neural networks must be modified to achieve the required smoothness. Determining the practicality of neural-network/nonlinear-inverse-dynamic controllers is an important area for future research.

ACKNOWLEDGMENTS

This research has been sponsored by the FAA and the NASA Langley Research Center under Grant No. NGL 31-001-252 and by the Army Research Office under Grant No. DAAL03-89-K-0092.

REFERENCES

1. M.G. Cox, "Practical Spline Approximation", in P.R. Turner, Ed., *Topics in Numerical Analysis, Lecture Notes in Mathematics*, Vol. 965, Springer-Verlag, Berlin, 1982, pp. 79-112.

2. S.H. Lane, "Theory and Development of Adaptive Flight Control Systems Using Nonlinear Inverse Dynamics," Ph.D. Dissertation, Princeton University, June, 1988.

3. J.S. Albus, "A New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC)," *Transactions of the ASME Journal of Dynamic Systems, Measurement, and Control*, Vol. 97, September 1975, pp. 220-227.

4. J.S. Albus, "Data Storage in the Cerebellar Model Articulation Controller (CMAC)," *Transactions of the ASME Journal of Dynamic Systems, Measurement, and Control*, Vol. 97, September 1975, pp. 228-233.

5. W.T. Miller, R.P. Hewes, F.H. Glanz, and L.G. Kraft, "Real-Time Dynamic Control of an Industrial Manipulator Using a Neural-Network-Based Learning Controller," *IEEE Transactions on Robotics and Automation*, Vol. 6, No. 1, February, 1990, pp. 1-9.

6. S.N. Singh and W.J. Rugh, "Decoupling in a Class of Nonlinear Systems by State Variable Feedback," *Transactions of the ASME, Journal of Dynamic Systems, Measurement, and Control*, Series G, Vol. 94, 1972, pp. 323-329.

7. A. Isidori, *Nonlinear Control Systems: An Introduction*, Springer-Verlag, Berlin, 1989.

8. S.N. Singh and A. Schy, "Output Feedback Nonlinear Decoupled Control Synthesis and Observer Design for Maneuvering Aircraft," *International Journal of Control*, Vol. 31, No. 4, 1980, pp. 781-806.

9. S.H. Lane and R.F. Stengel, "Flight Control Design Using Nonlinear Inverse Dynamics," *Automatica*, Vol. 24, No. 4, July, 1988, pp. 471-484.

10. R.F. Stengel, *Stochastic Optimal Control, Theory and Application*, John Wiley & Sons, Inc., New York, 1986.

11. H.L. Stalford, "High-Alpha Aerodynamic Model Identification of T-2C Aircraft Using the EBM Method," *Journal of Aircraft*, Vol. 18, 1981, pp. 801-809.

12. C. Fratter and R.F. Stengel, "Identification of Aerodynamic Coefficients Using Flight Testing Data," *AIAA Atmospheric Flight Mechanics Conference*, AIAA-83-2099, August, 1983.

13. M. Sri-Jayantha and R.F. Stengel, "Determination of Nonlinear Aerodynamic Coefficients Using the Estimation-Before-Modeling Method", *Journal of Aircraft*, Vol. 25, No. 9, September, 1988, pp. 796-804.

14. M. Sri-Jayantha and R.F. Stengel, "Data Acquisition System for High Angle of Attack Parameter Estimation," *Technical Soaring*, Volume IX, No. 4, pp. 85-95.

15. X. Silhouette and R.F. Stengel, "Estimation of the Aerodynamic Coefficients of the Navion Aircraft at High Angles of Attack and Sideslip," *AIAA Atmospheric Flight Mechanics Conference*, AIAA-87-2622, August, 1987, pp. 452-463.

16. G.J. Bierman, "Fixed Interval Smoothing with Discrete Measurements," *International Journal of Control*, Vol. 18, No. 1, January, 1973, pp. 65-75.

17. G. Rodriguez, "Kalman Filtering, Smoothing, and Recursive Robot Arm Forward and Inverse Dynamics," *Journal of Robotics and Automation*, Vol. RA-3, No. 6, December, 1987, pp. 624-639.

18. T. Poggio and F. Girosi, "Regularization Algorithms for Learning That Are Equivalent to Multilayer Networks," *Science*, Vol. 247, February 23, 1990, pp. 978-982.

19. D.J. Linse and R.F. Stengel, "Neural Networks for Function Approximation in Nonlinear Control," to be presented at the 1990 American Control Conference, San Diego, CA, May, 1990.

20. D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning Internal Representations by Error Propagation," *Parallel Distributed Processing: Explorations in the Microstructure of Cognitions, Volume 1: Foundations*, D.E. Rumelhart and J.L. McClelland, Eds., MIT Press, Cambridge, Massachusetts, 1986.

21. G. Cybenko, "Approximation by Superposition of a Sigmoidal Function," *Mathematics of Control, Signals, and Systems*, Vol. 2, No. 4, 1989, pp. 303-314.

22. D.A. Handelman, S.H. Lane, and J.J. Gelfand, "Characteristics of Integrated Rule-Based Systems and Neural Networks Enabling Robotic Skill Acquisition," *Proceedings of the BIRA Seminar on Neural Networks*, Ghent, Belgium, April 1989.

23. W.T. Miller, F.H. Glanz, and L.G. Kraft, "Application of a General Learning Algorithm to the Control of Robotic Manipulators," *International Journal of Robotics Research*, Vol. 6, No. 2, 1987, pp. 84-98.

24. L.G. Kraft and D.P. Campagna, "A Comparison of CMAC Neural Network and Traditional Adaptive Control," *Proceedings of the 1989 American Control Conference*, Vol. 1, 1989, pp. 884-889.

25. D.E. Knuth, *The Art of Computer Programming*, Vol. 3, Addison-Wesley Publishing Company, Reading, MA, 1973.

26. S.H. Lane, D.A. Handelman, and J.J. Gelfand, "Development of Adaptive B-Splines Using CMAC Neural Networks," *1989 IEEE INNS International Joint Conference on Neural Networks (IJCNN)*, Vol. 1, Washington, D.C., July 1989, pp. 683-688.

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CP-3095		

4. Title and Subtitle	5. Report Date
Joint University Program for Air Transportation Research 1989–1990	December 1990
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Frederick R. Morrell, Compiler	L-16848
9. Performing Organization Name and Address	10. Work Unit No.
NASA Langley Research Center Hampton, VA 23665-5225	505-66-01-02
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Aeronautics and Space Administration Washington, DC 20546 and Federal Aviation Administration Washington, DC 20546	Conference Publication
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This report summarizes the research conducted during the academic year 1989–1990 under the NASA/FAA sponsored Joint University Program for Air Transportation Research. The year-end review was held at Ohio University, Athens, Ohio, June 14–15, 1990. The Joint University Program is a coordinated set of three grants sponsored by NASA Langley Research Center and the Federal Aviation Administration, one each with the Massachusetts Institute of Technology (NGL-22-009-640), Ohio University (NGR-36-009-017), and Princeton University (NGL-31-001-252). Completed works, status reports and annotated bibliographies are presented for research topics, which include navigation, guidance and control theory and practice, aircraft performance, human factors, and expert systems concepts applied to airport operations. An overview of the year's activities for each university is also presented.

17. Key Words (Suggested by Authors(s))	18. Distribution Statement
Avionics Aircraft guidance, navigation, and control Human factors Expert systems	Unclassified—Unlimited Subject Category 01

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	197	A09

NASA FORM 1626 OCT 86

NASA-Langley, 1990

For sale by the National Technical Information Service, Springfield, Virginia 22161-2171